

W-MARKOV MEASURES, TRANSFER OPERATORS, WAVELETS AND MULTIREOLUTIONS

DANIEL ALPAY, PALLE JORGENSEN, AND IZCHAK LEWKOWICZ

ABSTRACT. In a general setting we solve the following inverse problem: Given a positive operators R , acting on measurable functions on a fixed measure space (X, \mathcal{B}_X) , we construct an associated Markov chain. Specifically, starting with a choice of R (the transfer operator), and a probability measure μ_0 on (X, \mathcal{B}_X) , we then build an associated Markov chain T_0, T_1, T_2, \dots , with these random variables (r.v) realized in a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and each r.v. taking values in X , and with T_0 having the probability μ_0 as law. We further show how spectral data for R , e.g., the presence of R -harmonic functions, propagate to the Markov chain. Conversely, in a general setting, we show that every Markov chain is determined by its transfer operator. In a range of examples we put this correspondence into practical terms: (i) iterated function systems (IFS), (ii) wavelet multiresolution constructions, and (iii) IFSs with random control. Our setting for IFSs is general as well: a fixed measure space (X, \mathcal{B}_X) and a system of mappings τ_i , each acting in (X, \mathcal{B}_X) , and each assigned a probability, say p_i which may or may not be a function of x . For standard IFSs, the p_i 's are constant, but for wavelet constructions, we have functions $p_i(x)$ reflecting the multi-band filters which make up the wavelet algorithm at hand. The sets $\tau_i(X)$ partition X , but they may have overlap, or not. For IFSs with random control, we show how the setting of transfer operators translates into explicit Markov moves: Starting with a point $x \in X$, the Markov move to the next point is in two steps, combined yielding the move from $T_0 = x$ to $T_1 = y$, and more generally from T_n to T_{n+1} . The initial point x will first move to one of the sets $\tau_i(X)$ with probability p_i , and once there, it will choose a definite position y (within $\tau_i(X)$), now governed by a fixed law (a given probability distribution). For Markov chains, the law is the same in each move from T_n to T_{n+1} .

CONTENTS

1. Introduction	2
2. General theory	4
2.1. Pairs of random variables and transfer operators	5
2.2. A formula for the conditional expectation	11
2.3. Markov processes	12
2.4. Discrete case	13
2.5. Martingales	14
2.6. Homogeneous Markov chains (HMC)	16
2.7. Multiresolutions and Cuntz-Krieger relations	20

1991 *Mathematics Subject Classification.* 37C30, 46L55, 47B65, 60J05, 60J10, 65T60.

Key words and phrases. Transfer operator, Markov chains, solenoid, wavelet multiresolution.

Acknowledgments: D. Alpay thanks the Earl Katz family for endowing the chair which supported his research. The second named author (PJ) wishes to thank the department of mathematics at Ben-Gurion University for hospitality during a 4 weeks research visit in the Spring of 2016, allowing for collaboration, and completion of this research.

2.8. The Schur algorithm and homogeneous Markov chains (HMC)	20
2.9. A summary of formulas	21
3. Solenoid probability spaces	22
3.1. Definitions	23
3.2. The multiresolution associated with a solenoid	29
3.3. Conditional expectations associated with a solenoid	33
3.4. A general setting and an inverse problem	33
4. Examples and applications: Transfer operators and Markov moves	34
4.1. First examples	35
4.2. Cases where σ is not onto	37
4.3. Solenoids associated with the unit circle	37
4.4. Fractal examples	45
4.5. The Gauss operator	46
References	47

1. INTRODUCTION

The purpose of our paper is to explore in two directions the interconnection between positive operators R defined in certain function spaces, on the one hand, and associated discrete time-random processes on the other. The direction back from R to the discrete time-random process, we refer to as the inverse problem. It includes the construction of the process itself. By contrast, the direct problem starts with a given discrete time-random process, and then computes the associated transfer operator, or sequence of transfer operators, and then finally uses the latter in order to determine properties of the given random process under consideration.

Our second purpose is a list of applications of our results in the general setting, the applications ranging from homogeneous Markov chains with white noise-input, dynamics of endomorphisms, including logistics maps, encoding mappings, invariant measures, wavelets in a general setting of multi-resolutions and associated transfer operators, also called Ruelle operators. In the case of a single positive operators R , we obtain, via a solution to the inverse problem, an associated generalized Markov processes, but its detailed properties will depend on a prescribed weight function W , hence the term W -Markov processes. In the case of a prescribed sequence of positive operators, we still obtain associated discrete time-random processes, now with each operator R_n accounting for the transfer of information from time n to time $n+1$. But these processes will not be Markov. Hence the Markov property is equivalent to $R_n = R$ for all n .

Returning to the case of our study of dynamics of endomorphisms, say σ in X , if the transfer operator R is σ -homogeneous, we show that the associated Markov processes will be of a special kind: when realized in the natural probability space of an associated solenoid $\text{Sol}_\sigma(X)$ (see Definition 3.7 for the latter), we arrive at multi-scale resolutions in $\mathbf{L}_2(\text{Sol}_\sigma(X), \mathcal{F}, \mathbb{P})$ (see Definition 3.16), with the scale of resolutions in question defined from the given endomorphism σ . In the case when σ is the scale endomorphism of a wavelet construction, we show that the wavelet multi-scale resolution will agree with that of the associated solenoid analysis. The latter framework is much more general, and covers a variety of multiresolution models.

TABLE 1. Increasing level of generality (each with its transfer operator and multiresolution; see Tables 2 and 4)

Case	$\mathbf{L}_2(\mathbb{R}, dx)$ \longrightarrow	$\mathbf{L}_2(\text{Sol}_\sigma(X), \mathbb{P})$ \longrightarrow	$\mathbf{L}_2(\Omega, \mathbb{P})$
------	---	---	------------------------------------

Before turning to the third theme in our paper, a few words on terminology: by a measure space (X, \mathcal{B}_X) we mean a set X and a sigma-algebra \mathcal{B}_X of subsets, each specified at the outset, usually with some additional technical restrictions. By a probability space, we mean a triple $(\Omega, \mathcal{F}, \mathbb{P})$, sample space Ω , sigma-algebra of events \mathcal{F} , and probability measure \mathbb{P} . We shall consider systems of random variables with values in measure spaces (X, \mathcal{B}_X) ; different random variables may take values in different measure spaces. Our first order of business is to show that for any pair of random variables, say A and B , each taking values in a measure space, there is an associated transfer operator R , depending only on A and B , which transfers information from one to the other. If A and B are independent, the associated operator R will be of rank-one, while if the sigma algebra generated by A is contained in that of B , then R will be the inclusion operator of the L_2 -spaces of the respective distributions, the distribution of A and that of B .

One source of motivation for our present work is a number of recent papers dealing with generalized wavelet multiresolutions, see e.g., [5, 32, 38, 39, 46, 53, 61], and harmonic analysis on groupoids. While these themes may seem disparate, they are connected via a set of questions in operator algebra theory; see e.g., [26, 43, 44]. The positive operators considered here are in a general measure theoretic setting, but we stress that there is also a rich theory of positive integral operators in the metric space setting, often called Mercer operators, and important in the approach of Smale and collaborators to learning theory, see e.g., [20, 59, 66]. However for our present use, the setting of the Mercer operators is too restrictive.

While various aspects of our settings may have appeared in special cases in anyone or the other of existing treatments of Markov chains, the level of generality, the questions addressed, and the specific and detailed interconnections, some surprising, revealed below, we believe have not. Relevant references include [18, 34, 36] and the papers cited therein.

Aside from the Introduction, the paper is divided into three sections. Since our approach to the applications involves some issues of a general nature, we found it best to begin with general theory, Section 2, covering a number of new results, all based on several intriguing operator theoretic features of general systems of random variables, and their associated transfer operators. This is developed first, and its relevance to discrete-time random processes is then covered in the remaining of Section 2. From there, we then turn to Markov chains, developed in this rather general and operator theoretic framework, and with an emphasis on transfer operator related issues. It is our hope that this will be of interest to readers both in operator theory, and in random dynamical systems and their harmonic analysis. We have thus postponed the applications to the last section. This is dictated in part by our focus on those Markov chains and associated dynamical systems which are induced by endomorphisms in measure spaces. In Section 3 we show that this setting can be realized in probability spaces over solenoids. Each endomorphism induces a solenoid, and a Markov chain of a special kind. The usefulness of this point of view is then

documented with a host of applications and detailed examples which we have included in several subsections in Section 4. We believe that our results in both the general theory and in our applications sections are of independent interest.

2. GENERAL THEORY

In this section, we consider the following general setting of random variables systems (r.v.s) on a prescribed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each r.v. taking values in a measure space (X, \mathcal{B}_X) ; different random variables may take values in different measure spaces. Our aim is to make precise transfer between the different r.v.s making up the system. For this purpose we concentrate on the case of a pair of r.v.s, say A and B . There is then an associated transfer operator $R = R_{A,B}$, depending only on A and B , which transfers information from one to the other. The transfer operator makes precise the intertwining of the two random variables. Indeed, if A and B are in fact given to be independent, then the associated operator R will be of rank-one, or zero in the case of zero means. On the other hand, if the sigma algebra generated by A is contained in that of B , then R will be the inclusion operator of the L_2 -spaces of the respective distributions, i.e., the distribution of A and that of B . We further show, in the general setting, that the product of the respective conditional expectations (the one for A and the one for B) are linked, via a factorization formula, by the transfer operator $R_{A,B}$. See Table 3 below.

While Section 2 is somewhat long and technical, it serves two important purposes: one, it offers lemmas to be used in the proofs of our main theorems later. The second purpose is to develop the tools we need in several inductive limit constructions to be used in our analysis of inverse problems, the inductive limits here concern the step of realizing infinite-dimensional discrete time-random processes as inductive limits of finite systems. For the finite systems themselves we develop here (the first five lemmas in Section 2) a new kernel analysis which will then be used later when we build the infinite dimensional probability models needed in the main theorems. As mentioned, a key tool is the notion of a transfer operator for a pair (or a finite number of) random variables. We shall include an analysis of the special case when one of the two r.v.s takes values in a discrete measure space. There are two reasons for this, one the interest in Markov chains with discrete state space, and the other is the study of such random variables as stopping time (see Definition 2.26).

Our approach to the analysis of finite systems of r.v.s is operator theoretic, relying on systems of isometries, co-isometries and projections, the latter in the form of conditional expectations. Of independent interest is our Corollary 2.36 which offers a representation of some operator relations known as the Cuntz-Krieger relations in operator algebra theory. Lemmas 2.2, 2.4, 2.9, 2.17, and 2.38 prepare the ground for what is to follow. Main results in the section includes Theorems 2.7, 2.19, 2.29, 2.30, and 2.39, as well as their corollaries and applications. Theorem 2.19 offers a model for the analysis of Markov processes in the general setting of our paper, Theorem 2.29 is a result which supplies a model for Markov chains driven by white noise. In this case we also compute an explicit invariant measure. This in turn is applied (Theorem 2.39) to a new random process realized naturally in a probability space over the Schur functions from complex analysis. Background references on calculus of random variables include [24, 36, 45, 51, 58, 60]; on classes of positive

operators (Ruelle operators) [12, 27, 41, 42]; and on algebras of operators in Hilbert space [6, 21, 22, 43, 48, 52, 53, 56, 62].

2.1. Pairs of random variables and transfer operators. Let (X, \mathcal{B}_X) be a measurable space. In this section, we define a transfer operator associated with two X -valued random variables, say A and B , defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution probability of A (also called “law”) is defined by

$$\mu_A(L) = \mathbb{P}(A^{-1}(L)), \quad L \in \mathcal{B}_X,$$

and so, with $\mathcal{M}(X, \mathcal{B}_X)$ denoting the space of real-valued measurable functions defined on X ,

$$\int_{\Omega} f(A(\omega)) d\mathbb{P}(\omega) = \int_X f(x) d\mu_A(x), \quad \forall f \in \mathcal{M}(X, \mathcal{B}_X),$$

(and similarly for B).

Definition 2.1. We denote by \mathcal{F}_A the sub sigma-algebra of \mathcal{F} defined by

$$(2.1) \quad \mathcal{F}_A = \{A^{-1}(L); L \in \mathcal{B}\}.$$

By definition of μ_A , and with \mathcal{F}_A introduced in Definition 2.1, the map

$$(2.2) \quad V_A f = f \circ A$$

is an isometry from $\mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$ onto $\mathbf{L}_2(\Omega, \mathcal{F}_A, \mathbb{P})$. For the adjoint operator V_A^* we have the following covariance (in a sense analogue to the one in mathematical physics and representation theory).

Lemma 2.2. *It holds that*

$$(2.3) \quad (V_A^* \psi)(x) = \mathbb{E}_{A=x}(\psi | \mathcal{F}_A), \quad \psi \in \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}).$$

Proof. We take $\psi \in \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ and $f \in \mathbf{L}_2(X, \mu_A)$. We have

$$\begin{aligned} \langle V_A^* \psi, f \rangle_{\mu_A} &= \langle \psi, V_A f \rangle_{\mathbb{P}} \\ &= \langle \psi, f \circ A \rangle_{\mathbb{P}} \\ &= \int_{\Omega} \psi(\omega) f(A(\omega)) d\mathbb{P}(\omega) \\ &= \int_{\Omega} f(A(\omega)) \mathbb{E}(\psi | \mathcal{F}_A) d\mathbb{P}(\omega). \end{aligned}$$

But \mathcal{F}_A is generated by the functions of the form

$$\chi_{A^{-1}(\Delta)} = \chi_{\Delta} \circ A, \quad \Delta \in \mathcal{B}_X,$$

and so there is a uniquely determined function $g \in \mathcal{M}(X, \mathcal{B}_X)$ such that $E(\psi | \mathcal{F}_A) = g \circ A$. (Uniqueness of g follows from the fact that $V_A : \mathbf{L}_2(X, \mathcal{B}_X, \mu_A) \rightarrow \mathbf{L}_2(\Omega, \mathcal{F}_A, \mathbb{P})$ is an isometry). Hence

$$\langle V_A^* \psi, f \rangle_{\mu_A} = \int_{\Omega} f(x) g(x) d\mu_A(x),$$

and hence the formula,

$$(V_A^* \psi)(x) = g(x) = \mathbb{E}_{A=x}(\psi \mid \mathcal{F}_A).$$

□

Corollary 2.3. *The measure $(\psi d\mathbb{P}) \circ A^{-1}$ is absolutely continuous with respect to μ_A , and for $\psi \in \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ we have*

$$(2.4) \quad (\psi d\mathbb{P}) \circ A^{-1} = g d\mu_A,$$

and

$$(2.5) \quad V_A^* \psi = \frac{(\psi d\mathbb{P}) \circ A^{-1}}{d\mu_A}.$$

Proof. From the previous proof we have on the one hand

$$\begin{aligned} \langle V_A^* \psi, f \rangle_{\mu_A} &= \int_{\Omega} f(A(\omega)) (\psi(\omega) d\mathbb{P}(\omega)) \\ &= \int_X f(x) ((\psi d\mathbb{P}) \circ A^{-1})(x) \end{aligned}$$

and on the other hand,

$$\begin{aligned} \langle V_A^* \psi, f \rangle_{\mu_A} &= \int_{\Omega} f(A(\omega)) (\psi(\omega) d\mathbb{P}(\omega)) \\ &= \int_{\Omega} f(x) g(x) d\mu_A(x) \\ &= \int_{\Omega} f(A(\omega)) g(A(\omega)) d\mathbb{P}(\omega), \end{aligned}$$

and the claim follows by comparing these two computations. □

With the above random variables A, B , we associate the positive operator $R_{A,B}$, which we call the transfer operator from A to B , defined by

$$(2.6) \quad R_{A,B} = V_A^* V_B,$$

see the figure below:

$$\begin{array}{ccc} \mathbf{L}_2(\mu_B) & \xrightarrow{R_{A,B}} & \mathbf{L}_2(\mu_A) \\ \searrow V_B & & \nearrow V_A^* \\ & \mathbf{L}_2(\Omega, \mathbb{P}) & \end{array}.$$

Note that both V_A^* and $R_{A,B}$ are positive operators in the following sense:

$$\psi \geq 0 \implies V_A^* \psi \geq 0$$

and

$$f \geq 0 \implies V_A^* V_B f \geq 0.$$

The following result shows that $R_{A,B}$ is a conditional expectation. In (2.8), by $\mathbb{E}(\cdot \mid \mathcal{F}_A)$ we mean the orthogonal projection of $\mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ onto $\mathbf{L}_2(\Omega, \mathcal{F}_A, \mathbb{P})$. It can also be defined as

$$(2.7) \quad \mathbb{E}(\psi \mid \mathcal{F}_A) = \frac{d(\psi d\mathbb{P})}{d\mathbb{P}_A}$$

in terms of Radon-Nikodym derivatives; see [57].

Lemma 2.4. *We have:*

$$(2.8) \quad \mathbb{E}(f \circ A \mid \mathcal{F}_B) = (R_{A,B}^* f) \circ B = (R_{B,A} f) \circ B, \quad f \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$$

$$(2.9) \quad \mathbb{E}(g \circ B \mid \mathcal{F}_A) = (R_{A,B} g) \circ A = (R_{B,A}^* g) \circ A, \quad g \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_B).$$

Proof. We prove (2.9). The proof of (2.8) is similar and follows from $(V_A^* V_B)^* = V_B^* V_A$. Let $f_1 \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_B)$, and $f_2 \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$. On the one hand, we have

$$\langle V_A^* V_B f_1, f_2 \rangle_{\mu_A} = \int_{\Omega} ((R_{A,B} f_1) \circ A)(\omega) (f_2 \circ A)(\omega) d\mathbb{P}(\omega).$$

On the other hand,

$$\begin{aligned} \langle V_A^* V_B f_1, f_2 \rangle_{\mu_A} &= \langle V_B f_1, V_A f_2 \rangle_{\mathbb{P}} \\ &= \int_{\Omega} (f_1 \circ B)(\omega) (f_2 \circ A)(\omega) d\mathbb{P}(\omega) \\ &= \int_{\Omega} (\mathbb{E}(f_1 \circ B \mid \mathcal{F}_A))(\omega) (f_2 \circ A)(\omega) d\mathbb{P}(\omega) \end{aligned}$$

by definition of the conditional expectation, and the result follows. \square

Corollary 2.5. *Let A, B and C be three random variables with transfer functions*

$$R_{A,B} : \mathbf{L}_2(X, \mathcal{B}_X, \mu_B) \longrightarrow \mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$$

and

$$R_{B,C} : \mathbf{L}_2(X, \mathcal{B}_X, \mu_C) \longrightarrow \mathbf{L}_2(X, \mathcal{B}_X, \mu_B).$$

Then the following chain rule holds for all $f \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_C)$ and $x \in X$:

$$(2.10) \quad (R_{A,B} R_{B,C} f)(x) = \mathbb{E}_{A=x}((R_{B,C}(f)) \circ B \mid \mathcal{F}_A).$$

Proof. We have

$$\begin{aligned} (R_{A,B} R_{B,C} f)(x) &= (V_A^* V_B V_B^* V_C f)(x) \\ &= (V_A^* \mathbb{E}_B V_C f)(x) \\ &= (V_A^* \mathbb{E}_B (f \circ C \mid \mathcal{F}_B))(x) \\ &= (V_A^* ((R_{B,C}(f)) \circ B))(x), \end{aligned}$$

and the result follows from Lemma 2.2. \square

In the following lemma, X is assumed locally compact, and $C_c(X)$ denotes the space of continuous functions on X with compact support.

Lemma 2.6. *Assume that X is a locally compact topological space, and that \mathcal{B} is the associated Borel sigma-algebra. Assume moreover that $R_{A,B}$ sends $C_c(X)$ into $C(X)$. Then it holds that*

$$(2.11) \quad \mathbb{E}(f \circ B \mid A = x) = (R_{A,B}(f))(x).$$

Proof. We denote by $\mathcal{F}_{A,x}$ the sigma-algebra generated by the set $\{A = x\}$. We have

$$\mathbb{E}(f \circ B \mid \mathcal{F}_{A,x}) = \mathbb{E}(f \circ B \mid \mathcal{F}_A) \mid \mathcal{F}_{A,x}.$$

Using the previous lemma, we can then write

$$\begin{aligned} \int_{A=x} (R_{A,B}f \circ A)(\omega) d\mathbb{P}(\omega) &= \int_{A=x} \mathbb{E}(f \circ B \mid \mathcal{F}_{A,x}) d\mathbb{P}(\omega), \quad \text{and} \\ \int_{A \neq x} (R_{A,B}f \circ A)(\omega) d\mathbb{P}(\omega) &= \int_{A \neq x} \mathbb{E}(f \circ B \mid \mathcal{F}_{A,x}) d\mathbb{P}(\omega), \end{aligned}$$

from which we get (2.11). \square

Theorem 2.7. *Let the following be as above: The probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the random variables A and B , and the respective measures μ_A and μ_B . Let also $R_{A,B}$ be the corresponding transfer operator. Then the following are equivalent:*

(i) $\mu_B \ll \mu_A$, with $W = \frac{d\mu_B}{d\mu_A}$

(ii) *It holds that*

$$(2.12) \quad \int_X R_{A,B}(f)(x) d\mu_A(x) = \int_X f(x) W(x) d\mu_A(x),$$

that is, $\frac{d\mu_A R_{A,B}}{d\mu_A} = W$.

Proof. Let $f_1 \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$ and $f_2 \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_B)$. From the proof of Lemma 2.6 we have:

$$\int_X f_1(x) (R_{A,B}f_2)(x) d\mu_A(x) = \int_{\Omega} (f_1 \circ A)(\omega) (f_2 \circ B)(\omega) d\mathbb{P}(\omega)$$

Setting $f_1(x) \equiv 1$, we obtain

$$\int_X (R_{A,B}f_2)(x) d\mu_A(x) = \int_X f_2(x) d\mu_B(x),$$

so that $d(\mu_A R_{A,B}) = d\mu_B$. By definition of W , we obtain (2.12). The converse is clear. \square

One can associate with the transfer operator $R_{A,B}$ two extreme cases: On the one end, if $\text{rank } R_{A,B} = 1$, this corresponds to having A and B independent, see Proposition 2.8. No information is passed from A to B . On the other end, if $R_{A,B} = I$, it corresponds to the sub-Markovian case.

Proposition 2.8. *The random variables A and B are independent if and only if the transfer operator $R_{A,B}$ has rank 1.*

Proof. Indeed, assume first A and B independent, and let $f \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$ and $g \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_B)$. We have:

$$\begin{aligned} \langle f, R_{A,B}g \rangle_{\mu_A} &= \langle V_A f, V_B g \rangle_{\mathbb{P}} \\ &= \int_{\Omega} ((f \circ A)(\omega))((g \circ B)(\omega)) d\mathbb{P}(\omega) \\ &= \left(\int_{\Omega} (f \circ A)(\omega) d\mathbb{P}(\omega) \right) \left(\int_{\Omega} (g \circ B)(\omega) d\mathbb{P}(\omega) \right) \\ &= \left(\int_X f(x) d\mu_A(x) \right) \left(\int_X g(x) d\mu_B(x) \right) \\ &= \langle f, 1 \rangle_{\mu_A} \langle 1, g \rangle_{\mu_B}, \end{aligned}$$

the product of the means of the respective random variables $f(A)$ and $g(B)$, and hence $R_{A,B}$ has rank one. In Dirac's notation, (ket-bra) we can write

$$R_{A,B} = |1\rangle_{\mu_A} \langle 1|_{\mu_B}.$$

□

Given two projections P_1 and P_2 on a Hilbert space, we recall (see [6, p. 376]) that the sequence $(P_2 P_1)^m$ converges strongly to the projection on the intersection of the corresponding spaces. Applied to $P_1 = \mathbb{E}_B$ and $P_2 = \mathbb{E}_A$ we obtain that $\lim_{m \rightarrow \infty} (\mathbb{E}_A \mathbb{E}_B)^m$ is the projection onto $\mathbb{E}_A(\mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P})) \cap \mathbb{E}_B(\mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}))$, that is the orthogonal projection onto $\mathbf{L}_2(\Omega, \mathcal{F}_A \cap \mathcal{F}_B, \mathbb{P})$.

Here we have a more precise formula:

Lemma 2.9. *With A, B and $\mathbb{E}_A, \mathbb{E}_B$ as above, let P denote the orthogonal projection onto the eigenspace corresponding to the eigenvalue 1 of $\lim_{m \rightarrow \infty} (\mathbb{E}_A \mathbb{E}_B)^m$. Then,*

$$(2.13) \quad \lim_{m \rightarrow \infty} (\mathbb{E}_A \mathbb{E}_B)^m \psi = \mathbb{E}(\psi | \mathcal{F}_A \cap \mathcal{F}_B) = V_A P R_{A,B} V_B^* \psi, \quad \forall \psi \in \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}).$$

Proof. The proof follows from the formula

$$(2.14) \quad (\mathbb{E}_A \mathbb{E}_B)^{m+1} = V_A (R_{A,B} R_{A,B}^*)^m R_{A,B} V_B^*, \quad m = 0, 1, \dots$$

which is true for $m = 0$ and proved by induction as follows:

$$\begin{aligned} (\mathbb{E}_A \mathbb{E}_B)^{m+1} &= (\mathbb{E}_A \mathbb{E}_B)^m V_A R_{A,B} V_B^* \\ &= \overbrace{(V_A (R_{A,B} R_{A,B}^*)^{m-1} R_{A,B} V_B^*)}^{\text{induction at rank } m} \overbrace{(V_A R_{A,B} V_B^*)}^{\mathbb{E}_A \mathbb{E}_B} \\ &= V_A (R_{A,B} R_{A,B}^*)^{m-1} R_{A,B} \underbrace{V_B^* V_A}_{R_{A,B}^*} R_{A,B} V_B^* \\ &= V_A (R_{A,B} R_{A,B}^*)^{m-1} R_{A,B} R_{A,B}^* R_{A,B} V_B^* \\ &= V_A (R_{A,B} R_{A,B}^*)^m R_{A,B} V_B^*. \end{aligned}$$

To conclude we remark that $\lim_{m \rightarrow \infty} (\mathbb{E}_A \mathbb{E}_B)^m$, being a projection, has spectrum consisting of the eigenvalues 0 and 1. Indeed, let $S = R_{A,B} R_{A,B}^*$. By the assumptions, the

projection-valued spectral resolution $E^{(S)}$ of the self-adjoint operator S satisfies

$$S = \int_0^1 t E^{(S)}(dt),$$

and so $\lim_{m \rightarrow \infty} S^m = E^{(S)}(\{1\})$, where $E^{(S)}(\{1\})$ (denoted by P in (2.13)) is the spectral projection onto

$$\{f \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_A) : Sf = f\}.$$

As a result we get

$$\mathbb{E}(\cdot \mid \mathcal{F}_A \cap \mathcal{F}_B) = V_A E^{(1)} R_{A,B} V_B^*.$$

□

For a related result, see [62].

As a corollary we have (where here and in the sequel we denote by \mathbb{E}_A the conditional expectation onto \mathcal{F}_A):

Corollary 2.10. *In the notation of the previous proposition and of its proof, let $S = R_{A,B} R_{A,B}^*$, and let $f \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$ and $\psi = V_A f$. The following are equivalent:*

- (1) $Sf = f$, i.e., $E^{(S)}(\{1\})f = f$.
- (2) ψ satisfies $\mathbb{E}_A \mathbb{E}_B \psi = \psi$.
- (3) ψ satisfies $\mathbb{E}_B \mathbb{E}_A \psi = \psi$.
- (4) $\mathbb{E}(\psi \mid \mathcal{F}_A \cap \mathcal{F}_B) = \psi$.

Proof. If T is a contraction from a Hilbert space \mathcal{H} into itself and $T\psi = \psi$ for some $\psi \in \mathcal{H}$, then we also have $T^*\psi = \psi$. Indeed, using $T\psi = \psi$ we obtain

$$\|\psi - T^*\psi\|^2 = \|T^*\psi\|^2 - \|\psi\|^2,$$

which is negative since T^* is also a contraction. Hence $\|\psi - T^*\psi\| = 0$ and $T^*\psi = \psi$. The proof of the corollary follows then by applying the above fact to $T = \mathbb{E}_A \mathbb{E}_B$. □

Corollary 2.11. *In the notation of the previous proposition, the following are equivalent for pairs of random variables A and B :*

- (1) $\mathcal{F}_A \subset \mathcal{F}_B$, (that is containment of the sigma-algebras of subsets of Ω)
- (2) $\mathbb{E}_A(\mathbf{L}_2(\Omega, \mathbb{P})) \subset \mathbb{E}_B(\mathbf{L}_2(\Omega, \mathbb{P}))$.
- (3) $\mathbb{E}_A \mathbb{E}_B = \mathbb{E}_A$, or equivalently $\mathbb{E}_A \leq \mathbb{E}_B$, where \leq denotes the standard ordering of projections.
- (4) $\mathbb{E}_B \mathbb{E}_A = \mathbb{E}_A$, equivalently $\mathbb{E}_A \leq \mathbb{E}_B$.
- (5) $R_{A,B} V_B^* = V_A^*$
- (6) $V_B R_{B,A} = V_A$.

Proof. This is essentially from the above, but see also the arguments outlined in Table 3 below. □

2.2. A formula for the conditional expectation. We are in the setting of Section 2.1. Let A be a X -valued random variable. For $f \in \mathbf{L}_2(X, \mathcal{B}_X)$ we denote by $M_{f \circ A}$ the operator of multiplication by $f \circ A$, from $\mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ into itself. The space of all these operators when f runs through $\mathbf{L}^\infty(X, \mathcal{B}_X)$ is a commutative von Neumann algebra, denoted \mathcal{M}_A . By Stone's theorem (see [56]), there exists a \mathcal{M}_A -valued measure \mathcal{E}_A on (X, \mathcal{B}_X) such that

$$(2.15) \quad M_{f \circ A} = \int_X f(x) \mathcal{E}_A(dx).$$

For every $L \in \mathcal{B}_X$, the operator $\mathcal{E}(L) \in \mathcal{M}_A$, and so is of the form $f \circ A$ for some $f \in \mathbf{L}^\infty(X, \mathcal{B}_X)$, namely $f = \chi_L$. From the equality

$$(2.16) \quad \mathcal{E}_A(L) = \chi_{\{A \in L\}} = \chi_{A^{-1}(L)}, \quad L \in \mathcal{B}_X,$$

we shall use the notation (after identifying the function and the corresponding multiplier)

$$(2.17) \quad \mathcal{E}_A(dx) = M_{\chi_{\{A \in dx\}}} = \chi_{\{A \in dx\}}$$

and rewrite (2.15) as

$$(2.18) \quad f \circ A = \int_X f(x) \chi_{\{A \in dx\}}, \quad \text{or} \quad (f \circ A)(\omega) = \int_X f(x) \chi_{\{A(\omega) \in dx\}}.$$

Remark 2.12. While $\chi_{\{A \in dx\}}$ is a heuristic notation, we stress that it is made precise via the spectral theorem in the form (2.15), and also by the conclusion of the next theorem.

Theorem 2.13. *Let $f \in \mathbf{L}_2(X, \mathcal{B}_X), \mu_A$. Then,*

$$(2.19) \quad \int_X f(x)^2 d\mu_A(x) = \int_\Omega \left(\int_X f(x) \chi_{\{A(\omega) \in dx\}} \right)^2 d\mathbb{P}(\omega).$$

Proof. Consider finite partitions $\pi = \{L_i, i = 1, \dots, m\}$ of X into sets of \mathcal{B} such that $L_i \cap L_j = \emptyset$ for $i \neq j$, and for every i chose $x_i \in L_i$. Let $|\pi| = \max_{i=1, \dots, m} |\mu_A(L_i)|$. We obtain a filter of \mathcal{B} -partitions along which limits are taken. By definition of the integral with respect with a measure we have:

$$\int_X f(x) \chi_{\{A(\omega) \in dx\}} = \lim_{|\pi| \rightarrow 0} \sum_{i=1}^m f(x_i) \chi_{\{A(\omega) \in L_i\}}.$$

But

$$\begin{aligned} \mathbb{E} \left(\sum_{i=1}^m f(x_i) \chi_{\{A(\omega) \in L_i\}} \right)^2 &= \sum_{i=1}^m f(x_i)^2 \mu_A(L_i) \\ &\rightarrow \int_X f(x)^2 d\mu_A(x), \quad \text{as } |\pi| \rightarrow 0, \end{aligned}$$

and the result follows. \square

We now consider the case of a discrete random variable. We shall assume that $A : \Omega \rightarrow \mathbb{N}_0$. So, $X = \mathbb{N}_0$, and the space $\mathbf{L}_2(X, \mathcal{B}_X, \mu_A)$ is the Hilbert space of $\ell_2(\mu_A)$ real-valued sequences $(\xi_n)_{n \in \mathbb{N}_0}$ such that

$$(2.20) \quad \sum_{n=0}^{\infty} \xi_n^2 \mathbb{P}(\{A = n\}) < \infty.$$

We have

$$(2.21) \quad (M_{f \circ A} \psi)(\omega) = \sum_{n=0}^{\infty} \xi_n \chi_{\{A=n\}}(\omega) \psi(\omega), \quad \forall \psi \in \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}), \quad \forall f = (\xi_n)_{n \in \mathbb{N}_0} \in \ell_2(\mu_A),$$

and

$$(2.22) \quad \mathcal{E}_A(\{n\}) = M_{\chi_{\{A=n\}}}, \quad n = 0, 1, \dots$$

Theorem 2.14. *The following formulas hold:*

$$(2.23) \quad \mathbb{E}(\psi \mid \mathcal{F}_A)(\omega) = \begin{cases} \int_X (V_A^* \psi)(x) \chi_{\{A \in dx\}}(\omega) & (\text{continuous case}) \\ \sum_{k=0}^{\infty} \frac{1}{\mathbb{P}(\{A=k\})} \left(\int_{\{A=k\}} \psi(\omega) d\mathbb{P}(\omega) \right) \chi_{\{A=k\}}(\omega) & (\text{discrete case}). \end{cases}$$

2.3. Markov processes. We follow the notation of the previous section, but our starting point is now a sequence of X -valued random variables T_0, T_1, \dots defined on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Axioms 2.15. (i) Let $\mathcal{G}_n \subset \mathcal{F}$ be the smallest sigma-algebra for which the variables T_0, \dots, T_n are measurable. We have that $V_n^* V_{n+1}$ does not depend on n , and (see (2.8))

$$(2.24) \quad \mathbb{E}(f \circ T_{n+1} \mid \mathcal{G}_n) = \mathbb{E}(f \circ T_{n+1} \mid \mathcal{F}_n) = R(f) \circ T_n, \quad n = 0, 1, \dots$$

(ii) The measures μ_0 and μ_1 are equivalent.

We refer to (2.24) as the Markov property in the present setting.

Remark 2.16. If in the expression $R_{A,B}f = V_A^* V_B f$, to $A = T_n$ and $B = T_{n+1}$ and if moreover R_{T_{n+1}, T_n} is independent of n we get

$$\mathbb{E}(f \circ T_{n+1} \mid \mathcal{F}_n) = (R(f)) \circ T_n$$

as a special case of (2.8). Iterating we get

$$\mathbb{E}(f \circ T_{n+k} \mid \mathcal{F}_n) = (R^k(f)) \circ T_n.$$

Lemma 2.17. *Condition (ii) from Axioms 2.15 holds if and only if $\mu_0(\{x : W(x) = 0\}) = 0$.*

Proof. Since $\mu_1 \ll \mu_0$ we can write

$$(2.25) \quad \mu_1(\Delta) = \int_{\Delta} W(x) d\mu_0(x), \quad \forall \Delta \in \mathcal{B}_X,$$

where $W = \frac{d\mu_1}{d\mu_0}$. Let $\Delta_0 = \{x \in X : W(x) = 0\}$. By (2.25) we have $\mu_1(\Delta_0) = 0$. Assume that $\mu_0(\Delta_0) > 0$. Then $\mu_0 \ll \mu_1$ will not hold.

Conversely, if $\mu_0(\Delta_0) = 0$, then W^{-1} is well defined μ_0 a.e., and $\mu_0 \ll \mu_1$ with $\frac{d\mu_0}{d\mu_1} = \frac{1}{W}$. \square

Definition 2.18. Assume the previous axioms in force, and set $W = \frac{d\mu_1}{d\mu_0}$. The sequence T_0, T_1, \dots is called a W -Markov process.

Given $\prod_{n=0}^{\infty} X$, we denote by π_n the n -th coordinate function:

$$\pi_n(x_0, x_1, \dots) = x_n$$

sent work, this product is always endowed with the cylinder sigma-algebra \mathcal{C} .

Theorem 2.19. *Let $(\Omega, \mathcal{F}, \mathbb{P}, (T_n)_{n \in \mathbb{N}_0})$ satisfy axioms (i) and (ii) above. Then there is a probability measure \mathbb{P}^\times on the Cartesian product $\prod_0^\infty X$, and an isomorphism \widehat{T} between $(\Omega, \mathcal{F}, \mathbb{P}, (T_n)_{n \in \mathbb{N}_0})$ and $(\prod_0^\infty X, \mathcal{C}, \mathbb{P}^\times, (\pi_n)_{n \in \mathbb{N}_0})$, meaning that*

$$(2.26) \quad \pi_n \circ \widehat{T} = T_n, \quad n = 0, 1, \dots$$

Proof. We define

$$(2.27) \quad \widehat{T}(\omega) = (T_0(\omega), T_1(\omega), \dots)$$

and

$$(2.28) \quad \mathbb{P}^\times(\Delta) = \mathbb{P}\left(\widehat{T}^{-1}(\Delta)\right), \quad \forall \Delta \in \mathcal{C},$$

in other words, \mathbb{P}^\times is the distribution of \widehat{T} . □

2.4. Discrete case. We first compute the transfer operator (see (2.6)) for a pair of random variables A and B on $(\Omega, \mathcal{F}, \mathbb{P})$ when A is discrete. See Section 3.3 for the notation. We denote by δ_n the Dirac function on \mathbb{N}_0 , that is

$$\delta_n(m) = \begin{cases} 1, & \text{if } m = n, \\ 0, & \text{if } m \neq n. \end{cases}$$

Proposition 2.20. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and A, B as above. Then the transfer operator $R_{B,A} : \ell_2(\mu_A) \rightarrow \mathbf{L}_2(X, \mathcal{B}_X, \mu_B)$ is given by:*

$$(2.29) \quad (R_{B,A}(\delta_n))(x) = \mathbb{E}_{B=x}(\{A = n\} \mid \mathcal{F}_B).$$

Proof. The result is immediate from Lemmas 2.2 and 2.4. We get for $x \in X$ and $n \in \mathbb{N}_0$

$$\begin{aligned} ((V_B^* V_A) \delta_n)(x) &= (V_B^* (\chi_{\{A=n\}}))(x) \\ &= \mathbb{E}_{B=x}(\{A = n\} \mid \mathcal{F}_B) \end{aligned}$$

where we have identified the indicator function $\chi_{\{A=n\}}$ on Ω with the subset

$$\{A = n\} = \{\omega \in \Omega; A(\omega) = n\}.$$

□

Note that, since $R_{B,A}^* = R_{A,B}$, by Lemma 2.4 we get for $f \in \mathbf{L}_2(X, \mathcal{B}_X, \mu_B)$

$$(2.30) \quad (R_{A,B}(f))(n) = \int_X f(x) \mathbb{E}_{B=x}(\{A = n\} \mid \mathcal{F}_B) d\mu_B(x).$$

Proposition 2.21. *On $\mathbb{N}_0 \times \mathbb{N}_0$ we have the following positive definite kernel:*

$$(2.31) \quad k_B(n, m) = \langle R_{A,B} \delta_n, R_{A,B} \delta_m \rangle_{\mathbf{L}_2(X, \mathcal{B}_X, \mu_B)} = \int_\Omega \chi_{\{A=n\}}(\omega) \mathbb{E}_B(\chi_{\{A=m\}})(\omega) d\mathbb{P}(\omega).$$

Proof. This follows from (2.30) and from the formula $R_{B,A}^* R_{B,A} = V_A^* \mathbb{E}_B V_A$ □

The last result concerns the case where both A and B are discrete.

Proposition 2.22. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and A, B as above, and assume that both A and B have discrete laws, with μ_A and μ_B both supported on the same countable discrete set, say S . Denote by $\delta_j^{(B)}$ the Dirac function viewed as a vector in $\ell_2(\mu_B)$. Then*

$$(2.32) \quad \left(R_{A,B} \left(\delta_j^{(B)} \right) \right) (i) = \mathbb{P} (B = j \mid A = i) .$$

Proof. This is immediate from Lemma 2.6. □

TABLE 2. Markov chains

Case	General state space $\mathcal{M}(X, \mathcal{B}_X)$	Discrete state space
Transition	$\mathbb{E} \left(f \circ T_{n+1} \mid \bigvee_{j=0}^n \mathcal{F}_j \right) =$ $= \mathbb{E} \left(f \circ T_{n+1} \mid \mathcal{F}_n \right)$	$\mathbb{E} (T_{n+1} = j \mid T_0 = i_0, \dots, T_n = i_n) =$ $= \mathbb{E} (T_{n+1} = j \mid T_n = i_n)$
Transfer operator	$(Rf)(x) = \mathbb{E} (f \circ T_{n+1} \mid T_n = x)$ $\forall f \in \mathcal{M}(X, \mathcal{B}_X)$	$p_{i,j} = \mathbb{P} (T_{n+1} = j \mid T_n = i)$
Harmonic functions	$(Rh)(x) = h(x)$	$\sum_j p_{i,j} h_j = h_i$

2.5. Martingales.

Definition 2.23. Let T_0, T_1, \dots be a Markov chain, with each T_n taking values in the space (X, \mathcal{B}_X) . Let M_0, M_1, \dots be another X -valued random process. We say that $(M_n)_{n \in \mathbb{N}_0}$ is a martingale with respect to $(T_n)_{n \in \mathbb{N}_0}$ if the condition

$$(2.33) \quad \mathbb{E} \left(M_{n+k} \mid \bigvee_{j=0}^n \mathcal{F}_{T_j} \right) = M_n, \quad \forall n, k \in \mathbb{N}_0$$

holds.

Proposition 2.24. *Let (X, \mathcal{B}_X) be a measure space, and let T_0, T_1, \dots be a X -valued Markov chain, with transition operator R acting on $\mathcal{M}(X, \mathcal{B}_X)$, and let h positive on X and such that $Rh = h$. Then, the process $M_n = h \circ T_n$, $n = 0, 1, \dots$ is a martingale (see Definition 2.23 for the latter) with respect to $(T_n)_{n \in \mathbb{N}_0}$.*

Proof. Since $(T_n)_{n \in \mathbb{N}_0}$ is a Markov chain we have for every $f \in \mathcal{M}(X, \mathcal{B}_X)$

$$(2.34) \quad \mathbb{E} \left(f \circ T_{n+k} \mid \bigvee_{j=0}^n \mathcal{F}_j \right) = \mathbb{E} (f \circ T_{n+k} \mid \mathcal{F}_n) = (R^k(f)) \circ T_n,$$

(by Lemma 2.8 with $B = T_{n+k}$ and T_n instead of A ; see also (3.40) for the solenoid).

Setting $f = h$ in (2.34) we get

$$\begin{aligned}\mathbb{E} \left(M_{n+k} \mid \bigvee_{j=0}^n \mathcal{F}_j \right) &= (R^k(h)) \circ T_n \\ &= h \circ T_n \\ &= M_n,\end{aligned}$$

which is the desired conclusion. \square

Remark 2.25. The same argument will hold for a positive function f such that $Rf = \lambda f$ for some $\lambda \neq 0$. Then, $M_n^{(\lambda)} = \lambda^{-n} f \circ T_n$ is a martingale with respect to $(T_n)_{n \in \mathbb{N}_0}$.

Definition 2.26. (*stopping time*) A stopping time K for a random process $(T_n)_{n \in \mathbb{N}_0}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a random variable $K : \Omega \rightarrow \mathbb{N}_0$ such that

$$(2.35) \quad K^{-1}(\{n\}) \in \bigvee_{j=0}^n \mathcal{F}_{T_j}, \quad \forall n \in \mathbb{N}_0.$$

Applying our previous analysis to the pair $A = (T_0, \dots, T_n)$ (with associated space X^{n+1}) and $B = K$, we have the following stopping time formula:

$$(2.36) \quad (R_{\{T_0, \dots, T_n\}, K})(m) = \mathbb{E}_{\{K=m\}}(f(T_0, \dots, T_n) \mid \mathcal{F}_K), \quad n, m \in \mathbb{N}_0, \quad \forall f \in \mathbf{L}_2 \left(\left(\prod_{n=0}^n X \right), \mu_n \right),$$

where μ_n denotes the joint distribution of $\{T_0, \dots, T_n\}$.

Definition 2.27. Let (X, \mathcal{B}_X) and (Y, \mathcal{B}_Y) be two measure spaces, and let $F : X \times Y \rightarrow X$ be a measurable function, where $X \times Y$ as a measure space is given the product sigma-algebra. Let $(T_n)_{n \in \mathbb{N}_0}$ be a Markov chain with values in X , and let $(\psi_n)_{n \in \mathbb{N}_0}$ be a system of independent identically distributed (i.i.d.) random variables with values in Y . If

$$(2.37) \quad T_{n+1} = F(T_n, \psi_n), \quad n \in \mathbb{N}_0,$$

then one says that $(T_n)_{n \in \mathbb{N}_0}$ is a homogeneous Markov chain (HMC). In details, the requirement is that

$$T_{n+1}(\omega) = F(T_n(\omega), \psi_n(\omega)), \quad \omega \in \Omega,$$

where Ω refers to the sample space in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which realizes the two processes; see also Theorems 2.29 and 2.30 below.

Recursion (2.37) is a feedback loop in the language of (non-linear) system theory. See Figure 1 below, and see e.g. [33] for information on feedback. We plan to explore these connections in a future publication.

There are many applications of these Markov processes, including to control, see [34], to feedback, see [36], and to Monte Carlo simulation, see e.g., [18] and [60, §5.5].

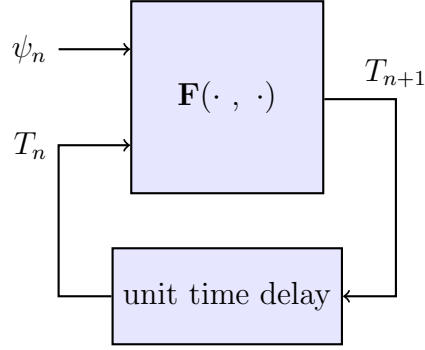


FIGURE 1. Feedback loop: illustration of the class of i.i.d. feedback processes from Definition 2.27.

2.6. Homogeneous Markov chains (HMC). As above (X, \mathcal{B}_X) is a set with a fixed sigma-algebra \mathcal{B}_X . We consider another measure-space (Y, \mathcal{D}) . Let ψ_0, ψ_1, \dots be a sequence of independent identically distributed (i.i.d.) Y -valued random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with probability distribution ν . Such a sequence is called a white noise or a driving sequence; see Remark 2.31. One way to construct such a sequence is as follows. One takes

$$\Omega_Y = \prod_{n=0}^{\infty} Y \quad (= Y^{\mathbb{N}_0})$$

endowed with the cylinder sigma-algebra \mathcal{C} (see for instance [45] for the latter), and the infinite product measure $\nu_{\infty} = \nu \times \nu \cdots$, and set

$$\psi_n(y_0, y_1, \dots) = y_n$$

Thus

$$(2.38) \quad \nu(D) = \nu_{\infty}(\psi_n^{-1}(D)), \quad D \in \mathcal{D}, \text{ and } n = 0, 1, \dots$$

In particular,

$$(2.39) \quad \int_{\Omega} F(\cdot, \psi_0(\omega)) d\nu_{\infty}(\omega) = \int_Y F(\cdot, y) d\nu(y)$$

We now consider a measurable map F from $X \times Y$ into X , where Y is another measure-space. We define

$$(2.40) \quad (R_F f)(x) = \int_Y f(F(x, y)) d\nu(y), \quad f \in \mathcal{M}(X, \mathcal{B}).$$

Let $\Omega_Y = \prod_{n=0}^{\infty} Y$, and for $\omega = (y_0, y_1, \dots) \in \Omega_Y$, we define (with $F_y = F(\cdot, y)$)

$$(2.41) \quad \begin{aligned} \omega|_n &= (y_0, \dots, y_n) \\ F_{\omega|_n} &= F_{y_n} F_{y_{n-1}} \cdots F_{y_1} F_{y_0}. \end{aligned}$$

We assume that

$$(2.42) \quad \bigcap_{n=1}^{\infty} F_{\omega|_n}(X) = \{x_{\omega}\}$$

is a singleton. We then set

$$(2.43) \quad V(\omega) = x_{\omega} \quad (\text{see (2.42)}).$$

Lemma 2.28. *Let $F : X \times Y \longrightarrow X$. The corresponding transfer operator in (2.40) is of the form*

$$(2.44) \quad (R_F f)(x) = \int_X f(t) \mu(dt, |x)$$

where $\mu(\cdot|x) \stackrel{\text{def.}}{=} (d\nu) \circ F_x^{-1}$, and where $F_x(\cdot) = F(x, \cdot) : Y \longrightarrow X$.

Proof. We have

$$\begin{aligned} \int_Y f(F(x, y)) d\nu(y) &= \int_Y (f \circ F_x)(y) d\nu(y) \\ &= \int_X f(t) (d\nu \circ F_x^{-1})(t). \end{aligned}$$

Hence $\mu(\cdot|x) = d\nu \circ F_x^{-1}$ as claimed. \square

Theorem 2.29. *Let ν be a probability measure on (Y, \mathcal{D}) , and let $\times_{n=0}^{\infty} \nu$ be the corresponding infinite product measure on Ω . Assume that (2.42) is in force, and let V be defined by (2.43). The formula*

$$(2.45) \quad \mu(B) = (\times_{n=0}^{\infty} \nu)(V^{-1}(B)), \quad B \in \mathcal{B}_X,$$

then defines a measure on (X, \mathcal{B}) which satisfies

$$(2.46) \quad \mu R_F = \mu,$$

that is

$$(2.47) \quad \iint_{X \times Y} f(F(x, y)) d\nu(y) d\mu(x) = \int_X f(x) d\mu(x), \quad \forall f \in \mathcal{M}(X, \mathcal{B}_X)$$

holds.

Proof. We define on Ω

$$\ell(y)(y_0, y_1, \dots) = (y, y_0, y_1, \dots), \quad \text{with } y \in Y \text{ and } \omega = (y_0, y_1, \dots) \in \Omega.$$

Then it is clear from (2.42) and (2.43) that

$$(2.48) \quad F_y V = V \ell(y),$$

since

$$F_y F_{\omega|n} = F_{\ell(\omega)|n+1}.$$

Note that (2.48) means that the following commutative diagram is in force:

$$\begin{array}{ccc} \Omega_Y & \xrightarrow{V} & X \\ \ell(y) \downarrow & & \downarrow F_y \\ \Omega_Y & \xrightarrow{V} & X \end{array}$$

We now prove (2.47). Let $f \in \mathcal{M}(X, \mathcal{B})$. Then R_F in (2.46) can be rewritten as

$$R_F f = \int_Y (f \circ F_y) d\nu(y).$$

With $\mathbb{Q} = \times_{n=0}^{\infty} \nu$ and $\mu = \mathbb{Q} \circ V^{-1}$ we get

$$\begin{aligned}
\int_X (R_F f)(x) d\mu(x) &= \int_X (R_F f)(x) (d\mathbb{Q} \circ V^{-1})(x) \\
&= \int_{\Omega_Y} ((R_F f) \circ V)(\omega) d\mathbb{Q}(\omega) \\
&= \int_{\Omega_Y} \int_Y ((f \circ F_y) \circ V)(\omega) d\mathbb{Q}(\omega) d\nu(y) \\
&\quad (\text{and using (2.48)}) \\
&= \int_{\Omega_Y} \int_Y (f \circ V \circ \ell(y)) d\mathbb{Q}(\omega) d\nu(y) \\
&= \int_{\Omega_Y} \int_Y (f \circ V)(\omega) d\mathbb{Q} \circ \ell(y)^{-1} d\nu(y) \\
&\quad (\text{and since } \mathbb{Q} \text{ is an infinite product measure}) \\
&= \int_{\Omega_Y} \int_Y (f \circ V)(\omega) d\mathbb{Q}(\omega) d\nu(y) \\
&= \int_X f(x) \int_Y (d\mathbb{Q} \circ V^{-1})(x) d\nu(y) \\
&= \int_X f(x) (d\mathbb{Q} \circ V^{-1})(x) \\
&\quad (\text{since } d\nu \text{ is a probability measure}) \\
&= \int_X f(x) d\mu(x) \quad (\text{with } \mu = \mathbb{Q} \circ V^{-1})
\end{aligned}$$

□

Theorem 2.30. *Let Y and F be as above. Let λ be a probability measure on (X, \mathcal{B}) , and let ψ_0, ψ_1, \dots be a sequence of i.i.d. Y -valued random variables with probability distribution ν . Then there exists a probability measure \mathbb{P} on (Ω_Y, \mathcal{C}) and a sequence of X -valued random variables T_0, T_1, \dots on Ω_Y such that:*

(1) λ is the distribution of T_0 , that is

$$(2.49) \quad \int_{\Omega} F(T_0(\omega), \cdot) d\mathbb{P}(\omega) = \int_X F(x, \cdot) d\lambda(x).$$

(2) We have

$$(2.50) \quad T_{n+1} = F(T_n, \psi_n), \quad n = 0, 1, \dots$$

(3) It holds that

$$(2.51) \quad \mathbb{E}(f \circ T_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(f \circ T_{n+1} \mid \mathcal{G}_n) = (R_F(f)) \circ T_n$$

where $\mathcal{F}_n = T_n^{-1}(\mathcal{B})$, where \mathcal{G}_n is the smallest sigma-algebra for which the variables T_0, \dots, T_n are measurable, and where R_F is given by (2.40).

(4) We have

$$(2.52) \quad \int_{\Omega_Y} (f_0 \circ T_0)(f_1 \circ T_1) \cdots (f_n \circ T_n) d\mathbb{P} = \int_X f_0(x) R_F(f_1 R_F(f_2 \cdots R_F(f_n h) \cdots)) d\lambda(x)$$

with $f_0, f_1, \dots, f_n \in \mathcal{M}(X, \mathcal{B})$.

Remark 2.31. Equation (2.50) is called an homogeneous Markov chain driven by white noise. The sequence ψ_0, ψ_1, \dots is called the driving sequence. See [18, p. 56]. For general background on time-homogeneous state equation and homogeneous Markov chains, [34, 36]. See also Theorem 3.9 below.

Remark 2.32. When R is not normalized one defines

$$R'(f) = \frac{R(fh)}{h}.$$

Then, $R'1 = 1$. The above construction applied to the pair $(R', h d\lambda)$ will lead to the same probability measure \mathbb{P} . This is because

$$\int_X f_0(x) R(f_1 R(f_2 \cdots R(f_n h) \cdots)) d\lambda(x) = \int_X f_0(x) R'(f_1 R'(f_2 \cdots R'(f_n) \cdots)) h(x) d\lambda(x).$$

Proof of Theorem 2.30. The proof is divided into three steps, which we outline.

STEP 1: Let $\omega = (y_0, y_1, \dots) \in \Omega_Y$ and define $\psi_n(\omega) = y_n$, and T_0, T_1, \dots via

$$(2.53) \quad T_n(\omega) = F(\cdots (F(F(F(T_0(\omega), y_0), y_1), \cdots), y_{n-1}), \dots)).$$

STEP 2: Formula (2.52) defines a unique probability measure \mathbb{P} on Ω_Y endowed with its cylinder sigma-algebra.

The existence of \mathbb{P} is an application of Kolmogorov's consistency principle; see for instance [51].

STEP 3: The above probability, and the random functions T_0, T_1, \dots have the desired properties.

See also Lemma 2.2 and 2.4. □

Corollary 2.33. The probability distribution of T_n is $\mu_n = \lambda R^n$, $n = 1, 2, \dots$

Proof. This follows from Theorem 2.30. See also Theorem 2.19. □

Corollary 2.34.

$$(2.54) \quad \mu_n(B) = (\lambda \times \underbrace{\nu \times \cdots \times \nu}_{n \text{ times}})(F_n^{-1}(B)), \quad B \in \mathcal{B}_X.$$

where

$$(2.55) \quad \begin{aligned} F_n(x, y_1, y_2, \dots, y_n) &= (F_{y_n} \cdots F_{y_1})(x) \\ &= F(\cdots F(F(F(x, y_1), y_2) \cdots, y_{n-1}), y_n). \end{aligned}$$

We set

$$\pi_1(x, y) = x.$$

Corollary 2.35. Let $B \in \mathcal{B}$ and $x \in X$. Then

$$(2.56) \quad \mathbb{E}(T_{n+1} \in B \mid T_n = x) = \nu(\pi_1^{-1}(x) \cap F^{-1}(B))$$

Proof. The result follows from Lemmas 2.2 and 2.4. \square

Background references on multiresolutions include [8, 10, 17, 23, 23].

2.7. Multiresolutions and Cuntz-Krieger relations. As a corollary of the previous analysis we now consider the case where possibly more than two random variables are given.

Corollary 2.36. *Given N random variables A_1, \dots, A_N with values in X , ($N = \infty$ is allowed), the following hold:*

$$(2.57) \quad V_{A_u}^* V_{A_v} = R_{u,v} \quad (\text{definition of the transfer operator from } A_u \text{ to } A_v)$$

$$(2.58) \quad V_{A_u}^* V_{A_u} = I_{\mathbf{L}_2(\mu_u)}, \quad u = 1, 2, \dots, N$$

$$(2.59) \quad V_{A_u} V_{A_u}^* = \mathbb{E}(\cdot | \mathcal{F}_{A_u}), \quad u = 1, 2, \dots, N$$

$$(2.60) \quad \sum_{u=1}^N V_{A_u} V_{A_u}^* = \mathbb{E}(\cdot | \cup_{u=1}^N \mathcal{F}_u),$$

$$(2.61) \quad \sum_{u=1}^N V_{A_u} V_{A_u}^* = \mathbb{E}(\cdot | \mathcal{F}) = I \quad \text{if } \mathcal{F} = \cup_{n=1}^N \mathcal{F}_{A_n}.$$

If $N = \infty$, the latter sums (2.60)-(2.61) converge in the strong operator topology.

Proof. This follows from Corollary 2.5 and Table 3. \square

Remark 2.37. Relations (2.57)-(2.61) can be seen as a generalization of the Cuntz-Krieger relations (see [21, 22] for the latter), and they lead to a multiresolution decomposition of the probability space $\mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P})$; see Section 4.3. A practical interpretation of formulae (2.60) and (2.61), is the assertion that certain random variables may be reconstructed by samples. In this case, the sampling is performed with the use of random variables as specified in the premise in Corollary 2.36. For a practical use of related sampling formulas in learning theory, see e.g., [66].

2.8. The Schur algorithm and homogeneous Markov chains (HMC). The Schur algorithm provides an application of the above analysis. We first recall the following (see [1, 11, 19, 29, 35, 65]). Let s be a function analytic and strictly contractive in the open unit disk \mathbb{D} (we will call such functions Schur functions, and denote their set by \mathcal{S}). Then, the functions s_1, \dots defined recursively by $s_0(z) = s(z)$ and

$$(2.62) \quad s_{n+1}(z) = \frac{s_n(z) - s_n(0)}{z(1 - \overline{s_n(0)}s_n(z))}, \quad n = 0, 1, \dots$$

belong to \mathcal{S} as long as $|s_n(0)| < 1$. The recursion stops at rank n if $|s_n(0)| = 1$. As already proved by Schur, this will happen if and only if s is a finite Blaschke product. The numbers $\rho_n = s_n(0)$, $n = 0, 1, \dots$ are called the Schur parameters of s , and determine uniquely the function s in terms of a partial fraction expansion

$$(2.63) \quad s(z) = \rho_0 + \frac{z(1 - |\rho_0|^2)}{1 - \overline{\rho_0}z} - \frac{1}{\rho_1 + \frac{z(1 - |\rho_1|^2)}{\overline{\rho_1} - \dots}}$$

See [68, p. 285]. When s is a finite Blaschke product, the sequence is finite, and its last element is of modulus 1.

Set $X = \mathcal{S} \setminus \{\text{unitary constants and finite Blaschke products}\}$ and $Y = \mathbb{D}$. We define

$$(2.64) \quad F(s, \rho)(z) = \frac{s(z) - \rho}{z(1 - s(z)\bar{\rho})}$$

which maps $X \times Y$ into X . We will also use the notations $F_\rho(s)$ and $(F_\rho(s))(z)$. We set $\Omega = \prod_{n=0}^{\infty} \mathbb{D}$ and,

$$\omega = (\rho_0, \rho_1, \dots) \quad \text{and} \quad \omega|_n = (\rho_0, \dots, \rho_n).$$

Furthermore, we define (see (2.41))

$$(F_{\omega|_n})(s) = (F_{\rho_n} F_{\rho_{n-1}} \cdots F_{\rho_1} F_{\rho_0})(s), \quad \text{see (2.63)}.$$

We denote by V the map

$$(2.65) \quad V(\omega) = s_\omega$$

where $s_\omega \in \mathcal{S}$ is uniquely defined element from ω via (2.63).

Lemma 2.38. *For every $\omega \in \Omega$ we have*

$$(2.66) \quad \bigcap_{n=0}^{\infty} F_{\omega|_n}(\mathcal{S}) = \{s_\omega\}$$

where $s_w = V(\omega)$, see (2.42).

Proof. This follows from the fact that a given Schur function is uniquely determined by the sequence of Schur coefficients when the latter is infinite. See [65]. \square

As a consequence of Theorem 2.29 we have the following result. In the proof the transfer operator now takes the form as in (2.40), that is

$$(2.67) \quad ((R_F f)(s))(z) = \int_{\mathbb{D}} f((F(s, \rho))(z)) d\nu(\rho).$$

Theorem 2.39. *Let ν be a probability measure on \mathbb{D} endowed with its Borel sigma-algebra, and let $\mathbb{Q} = \mathbb{Q}_\nu = \nu \times \nu \times \cdots$ be the corresponding infinite product measure on $\Omega = \prod_{n=0}^{\infty} \mathbb{D}$ endowed with the cylinder sigma-algebra. Then $\mu = \mathbb{Q}_\nu \circ V^{-1}$, where V is defined by (2.65), is a positive measure on \mathcal{S} (or, more precisely, on the set \mathcal{S} from which the unitary constants and finite Blaschke products have been removed) satisfying*

$$(2.68) \quad \mu R_F = \mu.$$

2.9. A summary of formulas. We now summarize some of the formulas obtained in this section, pertaining to two given X -valued random variables A and B

The proofs of these various formulas are given in the section. See in particular Lemmas 2.2, 2.4 and Corollary 2.5.

Remark 2.40. In the case when B is discrete, say $B : \Omega \rightarrow \mathbb{N}_0$, the formula for $R_{A,B}$ simplifies as follows:

$$(R_{A,B} f)(x) = \sum_{n=0}^{\infty} \frac{f(n)}{\mathbb{P}(\{B = n\})} \mathbb{E}_{\{A=x\}} (\chi_{\{B=n\}} \mid \mathcal{F}_A), \quad \forall x \in X$$

TABLE 3. Summary

Hilbert spaces	$\mathbf{L}_2(\mu_A) \longrightarrow \mathbf{L}_2(\Omega, \mathbb{P})$	$\mathbf{L}_2(\Omega, \mathbb{P}) \longrightarrow \mathbf{L}_2(\mu_A)$
Operators	$V_A f = f \circ A$	$(V_A^* \psi)(x) = \mathbb{E}_{A=x} (\psi \mid \mathcal{F}_A)$
Hilbert spaces	$\mathbf{L}_2(\Omega, \mathbb{P}) \longrightarrow \mathbf{L}_2(\Omega, \mathbb{P})$	$\mathbf{L}_2(\mu_B) \longrightarrow \mathbf{L}_2(\mu_A)$
Operators	$V_A V_B^* \psi = \mathbb{E} (\psi \mid \mathcal{F}_B \cap \{A = B\})$ (need $V_B^* \psi \in \mathbf{L}_2(\mu_A) \cap \mathbf{L}_2(\mu_B)$)	$(V_A^* V_B f)(x) = \mathbb{E}_{A=x} (f \circ B \mid \mathcal{F}_A)$ the tranfer operator $R_{A,B}$
Special case $A = B$	$V_A V_A^* = \mathbb{E} (\cdot \mid \mathcal{F}_A)$	$V_A^* V_A = I_{\mathbf{L}_2(\mu_A)}$
Product of the conditional expectations $\mathbb{E}_{\mathcal{F}_A}$ and $\mathbb{E}_{\mathcal{F}_B}$	$\mathbb{E}_{\mathcal{F}_A} \mathbb{E}_{\mathcal{F}_B} = V_A R_{A,B} V_B^*$	$\mathbf{L}_2(\Omega, \mathbb{P}) \longrightarrow \mathbf{L}_2(\mu_B) \longrightarrow$ $\longrightarrow \mathbf{L}_2(\mu_A) \longrightarrow \mathbf{L}_2(\Omega, \mathbb{P})$

for functions $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ such that

$$\sum_{n=0}^{\infty} |f(n)|^2 \mathbb{P}(\{B = n\}) < \infty, \quad i.e., f \in \ell_2(\mu_B).$$

3. SOLENOID PROBABILITY SPACES

Why the solenoids? A number of reasons. Given an endomorphism σ in a measure space, the associated solenoid Sol_σ is then a useful tool for the study of scales of multiresolutions (see Definitions 3.7 and 3.16). The latter includes those resolutions arising naturally from discrete wavelet algorithms, as well as from the study of non-reversible dynamics in ergodic theory in and physics. In fact it is not so much Sol_σ itself that is central in this program, but rather probability spaces $(\text{Sol}_\sigma, \mathcal{F}, \mathbb{P})$ where the solenoid is the sample space. It is the pair $(\mathcal{F}, \mathbb{P})$ which carries the information about the relevant scales of multiresolutions for the problem at hand, and the nature and the details of $(\mathcal{F}, \mathbb{P})$ change from one algorithm to the next; much like traditional wavelet analysis depend on scaling functions, father function, mother functions etc in $\mathbf{L}_2(\mathbb{R}^d)$. But the latter is too restrictive a framework; see e.g. Section 4 and [7, 17, 23]. See also Tables 1 and 4.

By “discrete wavelet algorithms” we mean recursive algorithms with selfsimilarity given by a scaling matrix. In one dimension, this may be just the N -adic scaling, but in general we allow for discrete time to be modelled by higher rank lattices, by more general discrete abelian groups, or even by infinite discrete sets with some given structure. For a given

time-series, even in this general form, we may always introduce an associated generating function. This will be a function in “dual frequency variables” in one or more complex variables, and called the frequency response function (see e.g., [17]). In many classical wavelet settings the given discrete wavelet algorithms may be realized in $\mathbf{L}_2(\mathbb{R}^d)$ for some d , but such a realization places very strong restrictions and limitations on the given multi-band filters making up the discrete wavelet algorithm at hand. We show that with the Hilbert space $\mathbf{L}_2(\text{Sol}_\sigma, \mathcal{F}, \mathbb{P})$, we can get around this difficulty, and still retain the useful features of multi-scale resolutions and selfsimilarity which makes the wavelet realizations so useful.

Motivated by multiresolutions in statistical computations, in many applications, and in particular in generalized wavelet algorithms, we study here a setting of dynamics of endomorphisms of measure spaces, denoting a given endomorphism by σ , say acting in X (see [12, 14, 40, 41, 42]). If the associated transfer operator R is further given to be σ -homogeneous (see Definition 3.1 below), we show that the associated R -Markov processes will be of a special kind: when realized in the natural probability space of an associated solenoid Sol_σ computed from σ , we then arrive at natural multi-scale resolutions inside the Hilbert space $\mathbf{L}_2(\text{Sol}_\sigma, \mathcal{F}, \mathbb{P})$, with the scale of resolutions in question defined from the given endomorphism σ . In the case when σ is the scale endomorphism of a wavelet construction, we show that the multi-scale resolution at hand will agree with that of the associated solenoid analysis. And when a wavelet is realizable in Euclidean space, for example on the real line \mathbb{R} , then we show that then \mathbb{R} is naturally embedded as a curve in the solenoid. Moreover, we identify the analogous multivariable setting with endomorphism and solenoid. Background references on analysis on solenoids and related multiresolutions include [10, 17, 27, 38, 42, 47].

In our discussion of solenoids and multiresolutions, we have here restricted the discussion to the commutative case, as our motivation is from stochastic processes. But in the recent literature, there is also an exciting, and somewhat parallel non-commutative theory of solenoids and their multiresolutions. It too is motivated (at least in part) by developments in the analysis of wavelet-multiresolutions, and the corresponding scaling operators. However, the relevant questions in the non-commutative theory are quite different from those addressed here. The relevant questions are simply different in the non-commutative theory. The differences between the two in fact reflect the dichotomy for two different notions of probability theory, the difference between (classical) commutative, versus non-commutative probability theory. Among the recent papers on the non-commutative theory, we mention [8, 9, 10, 47, 48], and the literature cited there.

3.1. Definitions. Consider a locally compact Hausdorff space X , with associated Borel sigma-algebra \mathcal{B} , let σ be a measurable endomorphism of X , which is onto. We denote by $\mathcal{M}(X, \mathcal{B})$ the space of all measurable functions from X into \mathbb{R} .

Definition 3.1. A map R from $\mathcal{M}(X, \mathcal{B})$ into itself is called a σ -transfer operator (or a Ruelle operator) if

$$(3.1) \quad Rf \geq 0, \quad \forall f \in \mathcal{M}(X, \mathcal{B}) \text{ satisfying } f(x) \geq 0, \quad \forall x \in X,$$

and if the pull-out property

$$(3.2) \quad R((f \circ \sigma)g) = fR(g), \quad \forall f, g \in \mathcal{M}(X, \mathcal{B})$$

holds.

As a first example we have:

Lemma 3.2. *Let (X, \mathcal{B}) be a measure-space and let A and B be two X -valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with transfer operator $R_{A,B}$ given by (2.6). Let σ be an endomorphism of X which is onto and such that*

$$(3.3) \quad \sigma \circ B = A.$$

Then $R_{A,B}$ satisfies the pull-out property (3.2). Moreover

$$(3.4) \quad R_{A,B}^* f = f \circ \sigma.$$

Proof. Indeed

$$\begin{aligned} (R_{A,B}((f \circ \sigma)g)) \circ A &= \mathbb{E} [((f \circ \sigma)g) \circ B \mid \mathcal{F}_A] \\ &= \mathbb{E} [(f \circ A)(g \circ B) \mid \mathcal{F}_A] \quad (\text{since } \sigma \circ B = A) \\ &= (f \circ A) \mathbb{E} [g \circ B \mid \mathcal{F}_A] \\ &= (f R_{A,B}(g)) \circ A. \end{aligned}$$

We now prove (3.4). We have

$$\begin{aligned} \langle g, f \circ \sigma \rangle_{\mu_B} &= \int_X g(x)((f \circ \sigma)(x)) d\mu_B(x) \\ &= \int_X R(g(f \circ \sigma))(x) d\mu_A(x) \\ &= \int_X f(x)(R(g)(x)) d\mu_A(x) \\ &= \langle R(g), f \rangle_{\mu_A}. \end{aligned}$$

□

We note that $\mathcal{F}_A \subset \mathcal{F}_B$ when (3.3) is in force. We now present an example of pairs of random variables for which neither $\mathcal{F}_A \subset \mathcal{F}_B$ nor $\mathcal{F}_B \subset \mathcal{F}_A$ hold. In particular they cannot be connected by an endomorphism of X .

Example 3.3. Consider the space $\{-1, 1\}$ with probability distribution

$$p(\{1\}) = p(\{-1\}) = \frac{1}{2}.$$

We take $\Omega = \prod_{n=1}^{\infty} \{-1, +1\}$, and \mathbb{P} the corresponding infinite product measure on the cylinder sigma-algebra. Let $a \in (0, 1)$ and define

$$(3.5) \quad E_a(\omega) = \sum_{k=1}^{\infty} \omega_k a^k,$$

where $\omega = (\omega_1, \omega_2, \dots) \in \Omega$ and thus $\omega_k \in \{-1, 1\}$. The random variable E_a takes values in \mathbb{R} , and its distribution, defined by

$$\alpha_a(x_1, x_2) = \mathbb{P}(\omega \in \Omega, x_1 < E_a(\omega) < x_2)$$

has Fourier transform

$$(3.6) \quad \widehat{\alpha}_a(t) = \prod_{k=1}^{\infty} \cos(a^k t).$$

It is known that (see [25, 27, 67]):

- (1) When $a < 1/2$ the distributions α_a are singular, and mutually singular.
- (2) When $a = 1/2$ we obtain the Lebesgue measure.
- (3) When $a \in (1/2, 1)$ the corresponding α_a are absolutely continuous with respect to Lebesgue measure, for almost all values of a . This is called the Erdős conjecture (see [30, 31]), and was proved in [67]. The only known value of $a > \frac{1}{2}$ for which α_a is known not to be absolutely continuous with respect to Lebesgue measure is the reciprocal of the golden ratio $a = \frac{\sqrt{5}-1}{2}$. See also [24, p. 48] for further references and information.

Taking a_1 and a_2 such that the corresponding distributions are mutually singular leads to random variables E_{a_1} and E_{a_2} which cannot be related by an endomorphism of X .

Definition 3.4. The solenoid $\text{Sol}_\sigma(X)$ associated with σ is the subset of sequences $(x_k)_{k \in \mathbb{N}_0}$ in $X^{\mathbb{N}_0}$ such that

$$(3.7) \quad \sigma(x_{k+1}) = x_k, \quad k = 0, 1, \dots$$

Remark 3.5. We think of a “point” in $\text{Sol}_\sigma(X)$ as a path-governed by σ , and hence $\text{Sol}_\sigma(X)$ as a path-space.

We set

$$(3.8) \quad \pi_k(x_0, x_1, \dots) = x_k, \quad k \in \mathbb{N}_0 \text{ and } (x_k)_{k \in \mathbb{N}_0} \in X^{\mathbb{N}_0},$$

and (3.7) can be rewritten as

$$(3.9) \quad \sigma \circ \pi_{n+1} = \pi_n, \quad n = 1, 2, \dots$$

The endomorphism σ is (in general) neither one-to-one nor onto. But:

Proposition 3.6. *The induced map $\hat{\sigma}$ defined by*

$$(3.10) \quad \hat{\sigma}(x_0, x_1, \dots) = (\sigma(x_0), x_0, x_1, \dots)$$

is one-to-one from $\text{Sol}_\sigma(X)$ onto itself, with inverse

$$(3.11) \quad \hat{\sigma}^{-1}(x_0, x_1, \dots) = (x_1, x_2, \dots)$$

Proof. One-to-oneness is clear. Let τ denote the map in (3.11). Then,

$$\hat{\sigma} \circ \tau(x_0, x_1, \dots) = \hat{\sigma}(x_1, \dots) = (\sigma(x_1), x_1, \dots) = (x_0, x_1, \dots)$$

since $\sigma(x_1) = x_0$ and

$$\tau \circ \hat{\sigma}(x_0, x_1, \dots) = \tau(\sigma(x_0), x_0, x_1, \dots) = (x_0, x_1, \dots).$$

□

We note that

$$(3.12) \quad \pi_0 \circ \hat{\sigma} = \sigma \circ \pi_0 \quad \text{and} \quad \pi_{n+1} \circ \hat{\sigma} = \pi_n, \quad n = 0, 1, \dots$$

Recall that the notation \mathcal{F}_A was introduced in (2.1). We set

$$\mathcal{F}_{\pi_n} = \mathcal{F}_n,$$

Definition 3.7. $\mathcal{F}_n = \pi_n^{-1}(\mathcal{B})$ is the sigma-algebra generated by the random variables $f \circ \pi_n$, where f runs through the measurable functions on (X, \mathcal{B}) .

As an immediate consequence of (3.12) we have:

Lemma 3.8. *In the notation of Definition 3.7, we have*

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \cdots$$

and $\bigcup_{n=0}^{\infty} \mathcal{F}_n = \mathcal{F}$.

Proof. Let $f \in \mathcal{M}(X, \mathcal{B}_X)$ and $n \in \mathbb{N}$. We have

$$f \circ \pi_n = (f \circ \sigma) \circ \pi_{n+1}$$

and so $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Since the sigma-algebra \mathcal{F} on $\text{Sol}_\sigma(X)$ is the cylinder sigma-algebra obtained from $\prod_{n=0}^{\infty} X$ we have that $\mathcal{F} = \bigvee_{n=0}^{\infty} \mathcal{F}_n$, where \bigvee denotes the lattice operation on sigma-algebras. \square

Theorem 3.9. *Let $(X, \mathcal{B}, \sigma, R, h, \lambda)$ be as above, and assume $Rh = h$. Then there exists a unique probability measure \mathbb{P} defined on the cylinder sigma-algebra on the associated solenoid $\text{Sol}_\sigma(X)$ such that*

$$(3.13) \quad \int_{\text{Sol}_\sigma(X)} (f_0 \pi_0)(\omega) (f_1 \pi_1)(\omega) \cdots (f_n \pi_n)(\omega) d\mathbb{P}(\omega) = \int_X (f_0(x) R(f_1 R(f_2 \cdots R(f_n h))))(x) d\lambda(x)$$

for all $n \in \mathbb{N}_0$ and $f_0, \dots, f_n \in \mathcal{M}(X, \mathcal{B})$. In the normalized case, we use $h \equiv 1$

Proof. We first remark that, in view of (3.9), \mathbb{P} (if it exists) is uniquely determined by

$$(3.14) \quad \int_{\text{Sol}_\sigma} (f \circ \pi_0)(\omega) d\mathbb{P}(\omega) = \int_X f(x) d\lambda(x),$$

and

$$(3.15) \quad \int_{\text{Sol}_\sigma(X)} (f \circ \pi_n)(\omega) d\mathbb{P}(\omega) = \int_X R^n(fh)(x) d\lambda(x).$$

For every $n \in \mathbb{N}_0$, there exists a measure \mathbb{P}_n on \mathcal{F}_n such that

$$(3.16) \quad \int_{\text{Sol}_\sigma(X)} (f \circ \pi_n)(\omega) d\mathbb{P}_n(\omega) = \int_X R^n(fh)(x) d\lambda(x).$$

Setting, in (3.15), $n+1$ instead of n , and taking into account (3.9), we have

$$\begin{aligned} \int_{\text{Sol}_\sigma(X)} (f \circ \pi_n)(\omega) d\mathbb{P}_{n+1}(\omega) &= \int_{\text{Sol}_\sigma(X)} (f \circ \pi_{n+1})(\omega) d\mathbb{P}_{n+1}(\omega) \\ &= \int_X R^{n+1}(f \circ \sigma)(x) d\lambda(x) \\ &= \int_X R^n(R(f \circ \sigma))(x) d\lambda(x) \\ &= \int_X (R^n f)(x) d\lambda(x) \end{aligned}$$

(by the pull-out property (3.2) since R is normalized)

$$= \int_{\text{Sol}_\sigma(X)} (f \circ \pi_n)(\omega) d\mathbb{P}_n(\omega).$$

By Kolmogorov's extension theorem (see e.g. [58]), the family $(\mathbb{P}_n)_{n \in \mathbb{N}}$ extends to a probability measure \mathbb{P} on the cylinder sigma-algebra. \square

We define the measure (λR) by

$$(3.17) \quad \int_X f(x) d(\lambda R)(x) = \int_X R(f)(x) d\lambda(x).$$

Definition 3.10. We say that σ is ergodic if

$$\cap_{n=1}^{\infty} \sigma^{-n}(\mathcal{B}_X) = \{\emptyset, X\},$$

modulo sets of λ -measure zero.

The examples of endomorphisms σ which we consider here are ergodic.

Theorem 3.11. *Let W be a positive measurable function on (X, \mathcal{B}) . The following are equivalent:*

(1) $\lambda R \ll \lambda$, and

$$(3.18) \quad \frac{d\lambda R}{d\lambda} = W.$$

(2) $\mathbb{P} \circ \hat{\sigma} \ll \mathbb{P}$ and

$$(3.19) \quad \frac{d\mathbb{P} \circ \hat{\sigma}}{d\mathbb{P}} = W \circ \pi_0$$

Proof. Assume that (1) is in force. To prove (2) we will show that (3.19) holds, or equivalently, that

$$(3.20) \quad \int_{\text{Sol}_\sigma(X)} \psi d\mathbb{P} = \int_{\text{Sol}_\sigma(X)} (\psi \circ \hat{\sigma})(W \circ \pi_0) d\mathbb{P}$$

for all measurable functions ψ on $\text{Sol}_\sigma(X)$. It suffices to take ψ of the form $\psi = f \circ \pi_n$ for $n = 0, 1, \dots$. We first consider the case $n = 0$. Let $\omega = (x_0, x_1, \dots)$. Then $\hat{\sigma}(\omega) = (\sigma(x_0), x_0, x_1, \dots)$, and so with $\psi = f \circ \pi_0$, the right-hand side of (3.20) is equal to

$$\begin{aligned} \int_{\text{Sol}_\sigma(X)} (f \circ \pi_0 \circ \hat{\sigma})(\omega)(W \circ \pi_0)(\omega) d\mathbb{P}(\omega) &= \int_{\text{Sol}_\sigma(X)} (f \circ \sigma)(x_0)(W \circ \pi_0)(\omega) d\mathbb{P}(\omega) \\ &= \int_{\text{Sol}_\sigma(X)} (f \circ \sigma \circ \pi_0)(\omega)(W \circ \pi_0)(\omega) d\mathbb{P}(\omega) \\ &= \int_{\text{Sol}_\sigma(X)} (((f \circ \sigma)W) \circ \pi_0)(\omega) d\mathbb{P}(\omega) \\ &= \int_X (f \circ \sigma)(x) W(x) d\lambda(x) \\ &= \int_X (R(f \circ \sigma))(x) d\lambda(x) \quad (\text{using (3.17) and (3.18)}) \\ &= \int_X f(x) d\lambda(x) \quad \text{since } R \text{ is normalized: } R1 = 1 \\ &\quad \text{and using the pullout property (3.2)} \\ &= \int_{\text{Sol}_\sigma(X)} (f \circ \pi_0)(\omega) d\mathbb{P}(\omega), \end{aligned}$$

which is the left-hand side of (3.20).

We now consider the case $n > 0$. The right-hand side of (3.20) is equal to:

$$\begin{aligned}
\int_{\text{Sol}_\sigma(X)} (f \circ \pi_n \circ \widehat{\sigma})(\omega)(W \circ \pi_0)(\omega) d\mathbb{P}(\omega) &= \int_{\text{Sol}_\sigma(X)} f(x_{n-1})(W \circ \pi_0)(\omega) d\mathbb{P}(\omega) \\
&= \int_{\text{Sol}_\sigma(X)} (f \circ \pi_{n-1})(\omega)(W \circ \pi_0)(\omega) d\mathbb{P}(\omega) \\
&= \int_X (R^{n-1}(f))(x) W(x) d\lambda(x) \\
&\quad \text{(where we have used (3.15))} \\
&= \int_X (R^n(f))(x) d\lambda(x) \\
&\quad \text{(by definition of } W) \\
&= \int_{\text{Sol}_\sigma(X)} (f \circ \pi_n)(\omega) d\mathbb{P}(\omega) \\
&\quad \text{(by (3.15))} \\
&= \int_{\text{Sol}_\sigma(X)} \psi(\omega) d\mathbb{P}(\omega),
\end{aligned}$$

by definition of $\psi = f \circ \pi_n$.

Conversely, we now assume that $\mathbb{P} \circ \widehat{\sigma} \ll \mathbb{P}$, with Radon-Nikodym derivative given by (3.19). Let $\psi = f \circ \pi_1$. We have:

$$(3.21) \quad \int_{\text{Sol}_\sigma(X)} (\psi \circ \widehat{\sigma})(\omega)(W \circ \pi_0)(\omega) d\mathbb{P}(\omega) = \int_{\text{Sol}_\sigma(X)} \psi(\omega) d\mathbb{P}(\omega).$$

Since $\pi \circ \widehat{\sigma} = \pi_0$, this latter equality is equivalent to:

$$\int_{\text{Sol}_\sigma(X)} ((fW) \circ \pi_0)(\omega) d\mathbb{P}(\omega) = \int_{\text{Sol}_\sigma(X)} (f \circ \pi_1)(\omega) d\mathbb{P}(\omega),$$

that is

$$(3.22) \quad \int_X (fW)(x) d\lambda(x) = \int_X (R(f))(x) d\lambda(x)$$

where we used (3.15). But (3.22) means that $\lambda R \ll \lambda$ with Radon-Nikodym derivative equal to W . \square

Remark 3.12. It follows from (3.15) that the probability distribution μ_n of the random variable π_n is equal to

$$d\mu_n(x) = W(x)((W \circ \sigma)(x)) \cdots ((W \circ \sigma^{n-1})(x)) d\lambda(x),$$

and that

$$(3.23) \quad \frac{d\mu_{n+1}}{d\mu_n} = (W \circ \sigma^n)(x).$$

We will assume that there the Radon-Nikodym derivative $W = \frac{d(\lambda R)}{d\lambda}$ exists, that is:

$$(3.24) \quad \int_X R(f)(x) d\lambda(x) = \int_X f(x) W(x) d\lambda(x), \quad \forall f \in \mathcal{M}(X, \mathcal{B}),$$

and that, furthermore, the Ruelle operator R in equations (4.13) and (3.1) is of Perron-Frobenius type in the sense that there exists $h \geq 0$, $h \in (X, \mathcal{B})$ such that

$$(3.25) \quad Rh = h,$$

and normalized to

$$(3.26) \quad \int_X h(x) d\lambda(x) = 1.$$

When R is not normalized one can replace R with the operator R' defined by

$$R'f = \frac{R(fh)}{h}.$$

It satisfies $R'1 = 1$. See Remark 2.32.

Definition 3.13. We will call $(X, \mathcal{B}, \sigma, R, h, \lambda)$ a generator for a path space when $\lambda R < \lambda$ and when R is normalized.

As a consequence of (3.9) we have (recall that σ is not one-to-one in general):

Proposition 3.14. *The distributions μ_k and μ_{k+1} are related by*

$$(3.27) \quad \mu_{k+1} \circ \sigma^{-1} = \mu_k,$$

meaning that

$$(3.28) \quad \mu_{k+1}(\sigma^{-1}(B)) = \mu_k(B), \quad \forall B \in \mathcal{B}.$$

Proof. By definition of μ_k we have:

$$\int_{\text{Sol}_\sigma(X)} (f \circ \pi_k)(\omega) d\mathbb{P}(\omega) = \int_X f(x) d\mu_k(x).$$

Hence,

$$\begin{aligned} \int_X (f \circ \sigma)(x) d\mu_{k+1}(x) &= \int_{\text{Sol}_\sigma(X)} ((f \circ \sigma) \circ \pi_{k+1})(\omega) d\mathbb{P}(\omega) \\ &= \int_{\text{Sol}_\sigma(X)} (f \circ \pi_k)(\omega) d\mathbb{P}(\omega) \quad (\text{using (3.9)}) \\ &= \int_X f(x) d\mu_k(x). \end{aligned}$$

It suffices to take $f(x) = \chi_B(x)$ to obtain (3.28). \square

Remark 3.15. (3.28) is independent of the given probability measure on the cylinder sigma-algebra.

3.2. The multiresolution associated with a solenoid. We begin with a table relative to the wavelet realization by unitary operators; the third column, related to the classical $\mathbf{L}_2(\mathbb{R}, dx)$ wavelets is elaborated upon in Section 4.3.

Definition 3.16. Let \mathcal{H} be a Hilbert space, let $U : \mathcal{H} \rightarrow \mathcal{H}$ be a unitary operator and let $(\mathcal{H}_n)_{n \in \mathbb{Z}}$ be an indexed family of closed subspaces such that:

- (i) $\mathcal{H}_{n+1} \subset \mathcal{H}_n$, $n \in \mathbb{Z}$,
 - (ii) $U^{-k}\mathcal{H}_0 = \mathcal{H}_k$, $k \in \mathbb{Z}$,
 - (iii) $\bigwedge_{k \in \mathbb{Z}} \mathcal{H}_k$ is at most one dimensional,
- and

(iv) $\bigvee_{k \in \mathbb{Z}} \mathcal{H}_k = \mathcal{H}$.

Here \bigwedge and \bigvee refer to the lattice operations applied to closed subspaces in \mathcal{H} .

When $(\mathcal{H}, U, (\mathcal{H}_n)_{n \in \mathbb{Z}})$ satisfy (i)-(iv), then we say that it is a multiresolution (or multi-scale resolution), and that U is the associated scaling operator.

Remark 3.17. Let U be a unitary operator which is part of a multiresolution, then it can be shown that the spectrum of U must be as follows: Except for the point $\lambda = 1$ occurring with at most multiplicity one, the spectrum of U must be absolutely continuous with uniform multiplicity infinity. This is an application of ideas of Wold, Lax-Phillips, and Stone-von Neumann; see [50, 52]. See also Remark 3.22 below.

We shall outline below a number of examples of multiresolutions, in wavelet theory and in dynamics more generally. This will make use of the theory we already developed in Section 2 above.

TABLE 4. Wavelets realization by unitary operators

The case	$\mathbf{L}_2(\mathbb{R}, dx)$	Fourier transform of $\mathbf{L}_2(\mathbb{R}, dx)$	General (solenoid) $\mathbf{L}_2(\text{Sol}_\sigma, \mathcal{C}, \mathbb{P})$
The unitary operator	$(Ug)(x) = \frac{1}{\sqrt{N}}g(x/N)$	$(U\gamma)(t) = \sqrt{N}\gamma(Nt)$	$(U\psi)(\omega) =$ $= (\psi \circ \widehat{\sigma})(m_0 \circ \pi_0)$
(1) Map onto the zero resolution subspace.	$K\xi = \sum_{k \in \mathbb{Z}} \xi_k \varphi(x - k)$ belongs to $\mathbf{L}_2(\mathbb{R}, dx)$ for $(\xi_n) \in \ell_2(\mathbb{Z})$.	$(Kf)(t) = f(t)\widehat{\varphi_0}(t)$ belongs to $\mathbf{L}_2(\mathbb{R}, dx)$ for $f \in \mathbf{L}_2(\mathbb{T})$.	$V_{\pi_0}f = f \circ \pi_0$ belongs to $\mathbf{L}_2(\text{Sol}_\sigma, \mathbb{P})$ for $f \in \mathbf{L}_2(X, \lambda)$.
(2) Average operator.	$(S\xi)_j = \sum_{k \in \mathbb{Z}} \xi_k a_{Nk-j}$	$(Sf)(t) = m_0(t)f(Nt)$	$Sf = m_0 \cdot f \circ \sigma$
Level zero resolution Invariant subspace for S (using (1) above)	$\xi \in \ell_2(\mathbb{Z}) \simeq \mathcal{H}_0$	$f \in \mathbf{L}_2(\mathbb{T}) \simeq \mathcal{H}_0$	$f \in \mathbf{L}_2(X, \lambda) \simeq \mathcal{H}_0^{\text{Sol}_\sigma}$

Remark 3.18. In the above table, in the second column, the map m_0 is a continuous function $m_0(z)$ on the unit circle, and the coefficients a_n are the Fourier coefficients of $|m_0(e^{it})|^2$. In the third column, we are in the special case where $W = |m_0|^2$, then W satisfies the conditions of Theorem 3.22.

Recall from Lemma 2.17 that if μ_0 and μ_1 are equivalent. Then, the set where W vanishes has measure zero.

Proposition 3.19. *Assume that μ_0 and μ_1 are equivalent, Then, the map*

$$(3.29) \quad U\psi = \sqrt{W \circ \pi_0} (\psi \circ \widehat{\sigma})$$

is unitary from $\mathbf{L}_2(\text{Sol}_\sigma(X), \mathbb{P})$ onto itself, and its inverse is given by

$$(3.30) \quad U^{-1}\psi = \frac{1}{\sqrt{W \circ \pi_1}} \psi \circ \widehat{\sigma}^{-1},$$

where $\widehat{\sigma}^{-1}$ is given by (3.11). We now check that $UU^{-1}\psi = \psi$. We have

$$\begin{aligned} UU^{-1}\psi &= \sqrt{W \circ \pi_0} (U^{-1}\psi) \circ \widehat{\sigma}^{-1} \\ &= \sqrt{W \circ \pi_0} \frac{1}{\sqrt{W \circ \pi_1 \circ \widehat{\sigma}}} \psi \circ \widehat{\sigma} \circ \widehat{\sigma}^{-1} \\ &= \psi \end{aligned}$$

since $\pi_1 \circ \widehat{\sigma} = \pi_0$. The proof that $U^{-1}U\psi = \psi$ is similar, and omitted.

Proof. The first claim is a consequence of (3.21) with $|\psi^2|$ instead of ψ . \square

Definition 3.20. Let

$$(3.31) \quad \mathcal{H}_0 = \{f \circ \pi_0 \mid f \in \mathbf{L}_2(X, h d\lambda)\},$$

the resolution subspace. The family $\mathcal{H}_n = U^{-n}\mathcal{H}_0$, $n \in \mathbb{Z}$, is called the multiresolution associated with the solenoid, and will be denoted by MR_σ .

Note that

$$(3.32) \quad \int_{\mathbf{L}_2(\mathbb{P})} |f \circ \pi_0|^2 d\mathbb{P} = \int_X |f(x)|^2 h(x) d\lambda(x).$$

Proposition 3.21. Let $f \in L^\infty(X)$. The multiplication map

$$(3.33) \quad M_{f \circ \pi_0}$$

sends $\mathbf{L}_2(\text{Sol}_\sigma, \mathbb{P})$ into itself, and \mathcal{H}_n into itself for all n . We have

$$\begin{array}{ccc} \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}) & \xrightarrow{U} & \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}) \\ M_{f \circ \pi_0} \downarrow & & \downarrow M_{f \circ \sigma \circ \pi_0} \\ \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}) & \xrightarrow{U} & \mathbf{L}_2(\Omega, \mathcal{F}, \mathbb{P}) \end{array}.$$

and the following covariance relation holds (see also Remarks 3.22 and 3.23)

$$(3.34) \quad UM_{f \circ \pi_0}U^{-1} = M_{f \circ \sigma \circ \pi_0}.$$

The map $f \mapsto M_f$ defines a representation of $L^\infty(X)$ by bounded operators on $\mathbf{L}_2(\text{Sol}_\sigma, \mathbb{P})$.

Proof. For $\psi \in \mathcal{M}(\Omega)$ we have

$$\begin{aligned} UM_{f \circ \pi_0}\psi &= \sqrt{W \circ \pi_0} (f \circ \pi_0 \circ \widehat{\sigma})(\psi \circ \widehat{\sigma}) \\ &= \sqrt{W \circ \pi_0} (f \circ \sigma \circ \pi_0)(\psi \circ \widehat{\sigma}) \quad (\text{where we use } \pi_0 \circ \widehat{\sigma} = \sigma \circ \pi_0; \text{ see (3.12)}) \\ &= M_{f \circ \sigma \circ \pi_0}U\psi. \end{aligned}$$

\square

Remark 3.22. Note that (3.34) is an instance of a covariance relation: It states that the representation M is unitarily equivalent to the representation obtained from it by substitution with the endomorphism σ . As a result, the projection valued measure determining M will satisfy the analogous covariance. This is outlined in (3.35) below. For the

convenience of the reader, let us give the following analogy: Consider the two canonical variables P and Q in the canonical commutation relation from quantum mechanics; in the Weyl exponentiated form. If E_Q denotes the projection valued spectral measure of Q , then the unitary one-parameter group $U(t)$, generated by P , satisfies a covariance in the form

$$U(t)E_Q(B)U(-t) = E_Q(B + t),$$

all for all Borel sets B , and all $t \in \mathbb{R}$. Here we use the word covariance in the same general context, but now for endomorphisms, also now instead for a single unitary operator. Many covariance relations have solutions that are unique up to unitary equivalence, for example the canonical $P - Q$ relation does; this is a form of the Stone-von Neumann uniqueness theorem. See [49, 50, 52].

Remark 3.23. The commutative von Neumann algebra \mathcal{M}_{π_0} of the multiplication operators $M_{f \circ \pi_0}$ with $f \in \mathbf{L}^\infty(X, \mathcal{B}_X)$ has the spectral representation (see Section 2.2 and equation (2.18))

$$M_{f \circ \pi_0} = \int_X f(x) \chi_{\{\pi_0 \in dx\}}$$

where $\chi_{\{\pi_0 \in dx\}}$ (also denoted by $\mathcal{E}_{\pi_0}(dx)$) is the projection-valued measure given by (2.16) and arising from the Stone theorem applied to \mathcal{M}_{π_0} ; see [56].

Define

$$\mathcal{E}_{\pi_0}^{(\sigma)}(\omega) = M_{\chi_{\{\pi_0 \in \sigma^{-1}(L)\}}}$$

As in (3.34) we arrive at the following selfsimilarity property for \mathcal{E}_{π_0} with U given by (3.29):

$$(3.35) \quad U \mathcal{E}_{\pi_0}(L) U^{-1} = \mathcal{E}_{\pi_0}(\sigma^{-1}(L)), \quad \forall L \in \mathcal{B}_X,$$

which we also rewrite as $U \mathcal{E}_{\pi_0} U^{-1} = \mathcal{E}_{\pi_0}^{(\sigma)}$.

We now give another interpretation of the resolution subspace \mathcal{H}_n . For \mathcal{F}_n , see Definition 3.7.

Proposition 3.24. *We have:*

$$(3.36) \quad \mathcal{H}_n = \mathbf{L}_2(\text{Sol}_\sigma, \mathcal{F}_n, \mathbb{P}) = \mathbb{E}_{\mathcal{F}_n}(\mathbf{L}_2(\text{Sol}_\sigma, \mathcal{F}, \mathbb{P})).$$

Proof. The proof follows from Corollary 2.11. □

Equation (3.37) below is a generalization of the classical notion of martingale.

Proposition 3.25. *Assume R normalized, i.e. $R1 = 1$. Then*

$$(3.37) \quad \mathbb{E}(f \circ \pi_{n+1} | \mathcal{F}_n) = R(f) \circ \pi_n, \quad n = 0, \dots$$

Proof. This follows from Lemma 2.4. □

Definition 3.26. The sequence $(T_n)_{n \in \mathbb{N}_0}$ of random variables from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into the measurable space (X, \mathcal{B}) is called a Markov chain if

$$(3.38) \quad \text{Pr}(T_{n+1} \in B | T_0, \dots, T_n) = \text{Pr}(T_{n+1} \in B | T_n) = (R(\chi_B)) \circ T_n$$

Proposition 3.27. *We have*

$$(3.39) \quad \mathbb{E}(f \circ \pi_{n+1} | \mathcal{F}_0, \dots, \mathcal{F}_n) = \mathbb{E}(f \circ \pi_{n+1} | \mathcal{F}_n), \quad n = 0, \dots$$

Proof. The result follows from Theorem 2.30. \square

Theorem 3.28. *Assume that R is normalized, and let $(\pi_n)_{n \in \mathbb{N}_0}$ be the stochastic process on $\text{Sol}_\sigma(X)$ defined by the coordinates. Then:*

$$(3.40) \quad \mathbb{E}(f \circ \pi_{n+k} | \mathcal{F}_n) = R^k(f) \circ \pi_n, \quad n = 0, \dots, \quad k = 1, 2, \dots$$

Proof. The proof uses the chain rule for conditional expectation and induction. It is enough to consider the case $k = 2$. We then have:

$$\begin{aligned} \mathbb{E}(f \circ \pi_{n+2} | \mathcal{F}_n) &= \mathbb{E}(\mathbb{E}(f \circ \pi_{n+2} | \mathcal{F}_{n+1}) | \mathcal{F}_n) \\ &= \mathbb{E}(R(f) \circ \pi_{n+1} | \mathcal{F}_n) \\ &= R^2(f) \circ \pi_n, \end{aligned}$$

where we used twice (3.37). \square

3.3. Conditional expectations associated with a solenoid.

Proposition 3.29. *The map $Vf = f \circ \pi_0$ is an isometry from $\mathbf{L}_2(X, h d\lambda)$ into $\mathbf{L}_2(\text{Sol}_\sigma, \mathbb{P})$.*

Proof. This is a corollary of Lemma 2.2. \square

Definition 3.30. The projection $\mathbb{E}_0 = VV^*$ in $\mathbf{L}_2(\text{Sol}_\sigma, \mathbb{P})$ is called the conditional expectation onto \mathcal{F}_0 of the multiresolution $(\text{Sol}_\sigma, \mathbb{P})$.

3.4. A general setting and an inverse problem. We now present a general setting, which includes the preceding analysis. We start from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a measurable space (X, \mathcal{B}) . We assume given a sequence of random variables $(T_n)_{n \in \mathbb{N}_0}$ from Ω to X , and an endomorphism σ from X into X . We assume that

$$(3.41) \quad \sigma \circ T_{n+1} = T_n, \quad n = 0, 1, \dots,$$

or, equivalently,

$$(3.42) \quad \forall \omega \in \Omega, \quad T_{n+1}(\omega) \in \sigma^{-1}(T_n(\omega)).$$

The map

$$(3.43) \quad \tilde{T}(\omega) = (T_0(\omega), T_1(\omega), \dots)$$

is measurable from Ω into $\text{Sol}_\sigma(X)$. It induces a probability measure \mathbb{P}^* on the cylinder sigma-algebra of $X^{\mathbb{N}}$ via the formula

$$(3.44) \quad \mathbb{P}^*(A) = \mathbb{P}(\tilde{T}^{-1}(A)).$$

The sequence $(T_n)_{n \in \mathbb{N}_0}$ generates a family of sigma-algebras, namely

$$(3.45) \quad \mathcal{F}_n = \{T_n^{-1}(A); A \in \mathcal{B}\}.$$

In view of (3.4) we have $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

We now recall a technical lemma, to be used in the proof of Theorem 3.32

Lemma 3.31. *Let (Ω, \mathcal{F}) and (X, \mathcal{B}) be two measurable spaces and let T be a map from Ω into X . Let*

$$\mathcal{F}_T = \{T^{-1}(B); B \in \mathcal{B}\}.$$

Then a real valued function ψ defined on Ω is \mathcal{F}_T -measurable if and only if it can be written in the form

$$\psi = f \circ T$$

for a uniquely defined \mathcal{B} -measurable function f .

Theorem 3.32. *There exists a positive operator defined on the space of measurable functions from X to \mathbb{R} such that*

$$(3.46) \quad \mathbb{E}(f \circ T_{n+1} | \mathcal{F}_n) = R(f) \circ T_n.$$

Proof. The existence of R follows from Lemma 3.31, and the positivity of R follows from the fact that a conditional expectation is an orthogonal projection. \square

Corollary 3.33.

$$(3.47) \quad R(1) = 1,$$

$$(3.48) \quad R((f \circ \sigma)g) = fR(g).$$

Proof. The first equation follows from setting $n = 0$ and $f \equiv 1$ in (3.46). The second equation is proved as follows. We have

$$\begin{aligned} \mathbb{E}(((f \circ \sigma)g) \circ T_1 | \mathcal{F}_0) &= \mathbb{E}((f \circ \sigma \circ T_1)(g \circ T_1) | \mathcal{F}_0) \\ &= \mathbb{E}((f \circ T_0)(g \circ T_1) | \mathcal{F}_0) \\ &= (f \circ T_0)\mathbb{E}(g \circ T_1 | \mathcal{F}_0) \\ &= (f \circ T_0)(R(g) \circ T_0) \\ &= (fR(g)) \circ T_0. \end{aligned}$$

\square

4. EXAMPLES AND APPLICATIONS: TRANSFER OPERATORS AND MARKOV MOVES

While in the abstract, as we showed, Markov chains are derived from positive operators R , acting on functions on a fixed measure space (X, \mathcal{B}_X) . Starting with a choice of R (the transfer operator), we then build a Markov chain T_0, T_1, T_2, \dots , with these random variables (r.v) realized in a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and each r.v. taking values in X , measurable of course with respect to the respective sigma algebras, \mathcal{F} on Ω , and \mathcal{B}_X on X . Conversely, every Markov chain is determined by its transfer operator.

The purpose of the examples below is to put this correspondence into more practical terms. The range of the examples we give will cover (i) iterated function systems (IFS), (ii) wavelet multiresolution constructions, and (iii) IFSs with random control.

An IFS on a fixed measure space (X, \mathcal{B}_X) is a system of mappings τ_i , each acting in (X, \mathcal{B}_X) , and each assigned a probability, say p_i which may or may not be a function of x . For standard IFSs it is not, but for wavelet constructions it is. In the latter, the functions $p_i(x)$ reflect the multi-band filters making up the wavelet algorithm. Moreover, the sets $\tau_i(X)$ partition X , but they may have overlap, or not. The Markov chains for the non-overlapping IFSs are simpler.

Returning to the general case, we now briefly sketch the idea behind Markov moves in IFSs with random control in a bit more detail. The examples below will supply hands-on cases, serving to illustrate the general idea.

The Markov move: Starting with a point x in X , the Markov move to the next point is in two steps, as follows, the combined two steps describing the move from T_0 to T_1 , and more generally from T_n to T_{n+1} . The initial point x will first move to one of the sets $\tau_i(X)$ with probability p_i , and once there, it will choose a definite position (within $\tau_i(X)$), and this second move will be prescribed by a fixed law (a given probability distribution); for

example, the law could be the uniform distribution, or something different. However, for Markov chains, the law is the same in the move from T_n to T_{n+1} , for all n .

4.1. First examples.

Example 4.1. In the first example, $X = [0, 1]$ and $\sigma(x) = 4x(1 - x)$, called the logistic map. von Neumann and Ulam proved that an invariant measure is

$$d\mu(x) = \frac{dx}{\pi\sqrt{x(1-x)}},$$

the Beta $B(\frac{1}{2}, \frac{1}{2})$ distribution, i.e. $\mu \circ \sigma^{-1} = \mu$. See [36, pp. 87-91]. The corresponding transfer operator is

$$(4.1) \quad (Rf)(x) = \frac{1}{2} \left(f\left(\frac{1 + \sqrt{1-x}}{2}\right) + f\left(\frac{1 - \sqrt{1-x}}{2}\right) \right)$$

We note that

$$\mu R \neq \mu.$$

We now turn to an example of a transfer operator $R_F : X \times Y \rightarrow X$

$$(R_F f)(x) = \int_Y f(F(x, y)) d\nu(y)$$

in which

$$(4.2) \quad \mu R_F = \mu$$

for the $B(\frac{1}{2}, \frac{1}{2})$ law μ . As a consequence of (4.2), we have that the corresponding probability measure \mathbb{P} in $(\prod_{n=0}^{\infty} X, \mathcal{F}, \mathbb{P})$ will be shift-invariant.

Example 4.2. We take $X = (0, 1)$. The endomorphism σ will depend on a parameter $u \in (0, 1)$, and is defined as follows. Set

$$(4.3) \quad \tau_0^{(u)}(x) = ux,$$

$$(4.4) \quad \tau_1^{(u)}(x) = u + (1 - u)x.$$

Then,

$$(4.5) \quad \sigma^{(u)}(x) = \begin{cases} \frac{x}{u}, & 0 < x \leq u, \\ -\frac{u}{1-u} + \frac{x}{1-u}, & u < x < 1. \end{cases}$$

Then,

$$\sigma^{(u)} \circ \tau_i^{(u)}(x) = x, \quad \text{for } i = 1, 2, \quad \text{and } x \in (0, 1).$$

Then,

$$(4.6) \quad R^{(u)} f(x) = \frac{1}{2} \left(f(\tau_0^{(u)}(x)) + f(\tau_1^{(u)}(x)) \right).$$

Let λ be the Lebesgue measure on $[0, 1]$. Then

$$(4.7) \quad d(\lambda R^{(u)}) = W^{(u)} d\lambda, \quad u \in (0, 1)$$

with

$$W^{(u)}(x) = \begin{cases} \frac{1}{2u}, & 0 \leq x < u, \\ \frac{1}{2(1-u)}, & u \leq x < 1. \end{cases}$$

Note that $W^{(u)}(x) \equiv x$ if and only if $u = \frac{1}{2}$. For every $u \in (0, 1)$ we have a quasi-invariant measure $\mathbb{P}^{(u)}$ such that

$$\frac{\mathbb{P}^{(u)} \circ \widehat{\sigma^{(u)}}}{d\mathbb{P}^{(u)}} = W^{(u)} \circ \pi_0.$$

Let $Y = \{0, 1\} \times (0, 1)$ and $d\nu = p_1 \times p_2$ be the product measure with $p_1(0) = p_1(1) = \frac{1}{2}$ and p_2 the uniform probability distribution on $(0, 1)$. Let furthermore

$$F(x, (i, u)) = \begin{cases} ux, & \text{if } i = 0, \\ (1 - u)x + u, & \text{if } i = 1, \end{cases}$$

and

$$\begin{aligned} (R_F f)(x) &= \int_Y f(F(x, (i, u))) d\nu(i, u) \\ (4.8) \quad &= \frac{1}{2} \int_0^1 (f(ux) + f((1 - u)x + u)) du \\ &= \frac{1}{2} \left(\frac{1}{x} \int_0^x f(t) dt + \frac{1}{1 - x} \int_x^1 f(t) dt \right). \end{aligned}$$

Now we show that the transfer operator which we just introduced has an invariant measure with absolutely continuous density.

Proposition 4.3. *Let R_F denote the transfer operator defined in (4.8), and set*

$$(4.9) \quad d\mu(x) = \frac{dx}{\pi \sqrt{x(1 - x)}}, \quad x \in (0, 1).$$

We then have

$$\mu R_F = \mu,$$

that is,

$$(4.10) \quad \int_0^1 (R_F f)(x) d\mu(x) = \int_0^1 f(x) dx, \quad \forall f \in \mathcal{M}((0, 1), \mathcal{B}).$$

Proof. For $d\mu(x) = G(x)dx$ to satisfy (4.10) we must have

$$G(y) = \frac{1}{2} \left(\int_y^1 \frac{G(x)}{x} dx + \int_0^y \frac{G(x)}{1 - x} dx \right).$$

Hence

$$\frac{G'(y)}{G(y)} = \frac{1}{2} \left(-\frac{1}{y} + \frac{1}{1 - y} \right),$$

and hence the result. □

Definition 4.4. We define the backward shift s on sequences of $\prod_{n=0}^\infty X$ by

$$(4.11) \quad s(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots).$$

Proposition 4.5. *In the setting of Theorem 2.29, let μ be an invariant measure for the transfer operator, and let π_0 be endowed with μ as probability law. Then the corresponding probability measure is shift-invariant:*

$$\mathbb{P} \circ s^{-1} = \mathbb{P}.$$

Proof. \mathbb{P} is built from the Kolmogorov construction by

$$(4.12) \quad \int_{\text{Sol}_\sigma(X)} (f_0\pi_0)(\omega)(f_1\pi_1)(\omega) \cdots (f_n\pi_n)(\omega) d\mathbb{P}(\omega) = \int_X (f_0(x)R(f_1R(f_2 \cdots R(f_nh))))(x) d\mu(x).$$

□

4.2. Cases where σ is not onto. When the endomorphism σ is not onto, the solenoid satisfies

$$\text{Sol}_\sigma(X) \subset \prod_{n=1}^{\infty} X_\infty^{(\sigma)}, \quad \text{where} \quad X_\infty^{(\sigma)} \stackrel{\text{def.}}{=} \cap_{n=1}^{\infty} \sigma^n(X),$$

and the latter can be a very small set, as we now illustrate.

Example 4.6. Take $X = [0, 1]$ and $\sigma(x) = 2x(1 - x)$. Then

$$\sigma(X) = [0, \frac{1}{2}] \quad \text{and} \quad X_\infty^{(\sigma)} = \left\{0, \frac{1}{2}\right\}.$$

The solenoid consists of the two points

$$(0, 0, \dots) \quad \text{and} \quad \left(\frac{1}{2}, \frac{1}{2}, \dots\right).$$

Example 4.7. This example is from complex dynamics. We take $X = \mathbb{C}$ and for a pre-assigned $c \in \mathbb{C}$,

$$\sigma_c(z) = z^2 + c.$$

Then $X_\infty^{(\sigma)}$ is the Julia set, see [54].

4.3. Solenoids associated with the unit circle. In the period since the mid 1990ties, the term wavelet has come to have a broader meaning: From referring to systems of bases in $\mathbf{L}_2(\mathbb{R})$ with dyadic scale symmetry, wavelet now typically refers to finite systems of functions on a suitable measure space that can be used in order to construct either an orthonormal basis, or frame basis by means of operators connected to algebraic and geometric information involving a notion of scaling function. The latter often in the form of a probability measure on a solenoid-measure space. In the case of fractals, there are natural choices of finite systems of functions yielding very well-behaved orthonormal bases, and thus giving direct information about the topological structure of the particular fractal involved. Our framework below makes use of solenoids (from endomorphisms) in order to offer an even more inclusive framework for wavelet bases and multiresolutions. Background references for the present section include [10, 13, 17, 27, 28, 38, 42, 43, 47].

4.3.1. Definition. Starting from a continuous function $m_0(z)$ on the unit circle \mathbb{T} and $N \in \{2, 3, \dots\}$ one can construct (at least) two representations of the algebra of operators generated by two operators T, U such that U is unitary and $UTU^{-1} = T^N$ (such an algebra is an algebra generated by a group of the kind studied in [13] by Baumslag and Solitar). To be more precise let $R = R_{m_0}$ denote the corresponding Ruelle operator:

$$(4.13) \quad (Rf)(z) = \frac{1}{N} \sum_{w^N=z} |m_0(w)|^2 f(w).$$

When $R1 = 1$ the infinite product $\prod_{u=1}^{\infty} \frac{m_0(e^{\frac{2\pi i t}{\sqrt{N}^u}})}{\sqrt{N}}$ belongs to $\mathbf{L}_2(\mathbb{R})$, and is the Fourier transform of the scaling function φ_0 . The space

$$(4.14) \quad \mathcal{H}_0 = \{\widehat{\varphi_0}(t)f(t); f \text{ function on } \mathbb{R}/\mathbb{Z} \text{ measurable and } f(t) = f(t+1)\}$$

is the 0-resolution subspace of the multiresolution

$$(4.15) \quad \mathcal{H}_k = \{2^{-k/2}\widehat{\varphi_0}(t/2^k)f(t); f \text{ measurable function on } \mathbb{R}/\mathbb{Z} \text{ i.e. } f(t) = f(t+1)\}, \quad k \in \mathbb{Z}.$$

One defines a representation ρ of $\mathbf{L}_{\infty}(\mathbb{T})$ into $\mathbf{B}(\mathbf{L}_2(\mathbb{R}))$ as follows: If $f \in \mathbf{L}_{\infty}(\mathbb{T})$ with associated Fourier series $f(e^{it}) = \sum_{n \in \mathbb{Z}} \widehat{f}(n)e^{int}$, one sets

$$(4.16) \quad \rho(f)(g) = \sum_{n \in \mathbb{Z}} \widehat{f}(n)g(x-n), \quad g \in \mathbf{L}_2(\mathbb{R}).$$

In this paper we remove the $\mathbf{L}_2(\mathbb{R}, dx)$ requirement (which we assumed in [2, 3]) from the wavelet setting. Now wavelet multiresolutions may be viewed as a special case of a probability space multiresolution. In the latter, the resolution subspaces will be specified by a system of conditional expectations. In the classical wavelet application, the solenoid becomes the real line, realized as a dense curve in $\text{Sol}_{\sigma}(X)$, and the solenoid measure P becomes Lebesgue measure.

We now consider the special case where X is equal to the unit circle $\mathbb{R}/\mathbb{Z} = \mathbb{T}$ and $\sigma(z) = z^N$. When using the notation $z = e^{2\pi i t}$ we have $\sigma(t) = Nt \pmod{1}$.

The solenoid G_N is a compact group, included in $\prod_{k=0}^{\infty} \mathbb{T}$, and consists of the sequences $z = (z_0, z_1, z_2, \dots) \in \prod_{k=0}^{\infty} \mathbb{T}$ such that

$$z_{k+1}^N = z_k, \quad k = 1, 2, \dots$$

See [42]. We define

$$(4.17) \quad \sigma(z_0, z_1, z_2, \dots) = (z_0^N, z_1^N, z_2^N, \dots) = (z_0^N, z_0, z_1, \dots)$$

and

$$(4.18) \quad \tau(z_0, z_1, z_2, \dots) = (z_1, z_2, \dots)$$

We have

$$\sigma \circ \tau = \tau \circ \sigma = I.$$

It is the dual of the discrete group $\mathbb{Z}[1/N]$, with characters $\chi\left(\frac{\ell}{N^k}\right)$ given by

$$(4.19) \quad \langle \chi\left(\frac{\ell}{N^k}\right), z \rangle = z_k^{\ell}, \quad k, \ell = 0, 1, \dots$$

See [14, 17]. Note that (4.19) is well defined since

$$\langle \chi\left(\frac{N\ell}{N^{k+1}}\right), z \rangle = z_{k+1}^{N\ell} = (z_{k+1}^N)^{\ell} = z_k^{\ell} = \langle \chi\left(\frac{\ell}{N^k}\right), z \rangle.$$

4.3.2. *Ruelle operators and wavelets.* We use the term Ruelle operator consistent with [4, 12, 17, 63] to indicate a transfer operator which governs branching in a number of different context. Every filter in the family we have can be realized as a wavelet filter on the solenoid. Fix a low-pass filter m_0 with the usual properties, and define

Two cases occur: When the function identically equal to 1 (denoted in this paper by $\mathbf{1}$) is an eigenvalue of R with eigenvalue 1, that is,

$$\frac{1}{N} \sum_{\substack{w \in \mathbb{T} \\ w^N = z}} |m_0(w)|^2 \equiv 1,$$

one can construct φ_0 and use the space $\mathbf{L}_2(\mathbb{R}, dx)$. We study the representations of the algebra generated by (U, T) such that

$$(4.20) \quad UTU^{-1} = T^N$$

We take

$$U \left(\sum \xi_n \varphi_0(\cdot - n) \right) = m_0(z) f(z^N) \quad \text{with} \quad f(z) = \sum \xi_n z^n.$$

Thus we have a slanted Toeplitz matrix

$$(S\xi)_n = \sum_{j \in \mathbb{Z}} a_{n-jN} \xi_j.$$

The following result reflects the scaling law for the father function φ_0 of the wavelet under consideration,

$$(4.21) \quad \varphi_0(x) = \sqrt{N} \sum_{k \in \mathbb{Z}} a_k \varphi_0(Nx - k), \quad x \in \mathbb{R},$$

where

$$(4.22) \quad m_0(x) = \sum_{k \in \mathbb{Z}} a_k e^{2\pi i k x}.$$

Lemma 4.8. *For the operators K and S (see (1) and (2) in Table 4 above) we define*

$$K : \ell_2(\mathbb{Z}) \longrightarrow \mathcal{H}_0 \text{ (the zero resolution subspace in } \mathbf{L}_2(\mathbb{R}) \text{)}$$

by

$$(K\xi)(x) = \sum_{n \in \mathbb{Z}} \xi_n \varphi_0(x - n).$$

Then

$$(4.23) \quad KS = UK$$

holds, that is the following diagram is commutative:

$$\begin{array}{ccc} \mathbf{L}_2(\mathbb{R}) & \xrightarrow{U} & \mathbf{L}_2(\mathbb{R}) \\ K \uparrow & & \uparrow K \\ \ell_2(\mathbb{Z}) & \xrightarrow{S} & \ell_2(\mathbb{Z}), \end{array}$$

where

$$(U\gamma)(x) = \frac{1}{\sqrt{N}} \gamma(x/N), \quad \gamma \in \mathbf{L}_2(\mathbb{R}, dx).$$

Proof. We have for $\xi \in \ell_2(\mathbb{Z})$:

$$(KS\xi)(x) = \sum_{n \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} a_{n-jN} \xi_j \varphi_0(x-n)$$

and

$$\begin{aligned} (UK\xi)(x) &= \sum_{j \in \mathbb{Z}} \xi_j \frac{1}{\sqrt{N}} \varphi_0\left(\frac{x-jN}{N}\right) \\ &= \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \xi_j a_k \varphi_0(x-jN-k) \\ &\text{and, with the change of variable } n = jN + k, \\ &= \sum_{n \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \xi_j a_{n-jN} \varphi_0(x-n), \end{aligned}$$

and the result follows. \square

More generally for many choices of filters (see (4.22)) there are no $\mathbf{L}_2(\mathbb{R}, dx)$ -solution to (4.21), and then one leaves the setting of $\mathbf{L}_2(\mathbb{R})$. We still get counterparts of (3.34) and (3.48) using the solenoid.

Proposition 4.9. *The operator R in (4.13) is bounded from $\mathbf{L}_2(\mathbb{T}, d\lambda)$ into itself, and its adjoint is given by the formula*

$$(4.24) \quad (R^*f)(z) = |m_0(z)|^2 f(z^N).$$

We now discuss the multiresolution associated with m_0 and its relationships with the multiresolution MR_σ . We first note that the space \mathcal{H}_0 defined by (4.14) is equal to the closed linear span of the functions $x \mapsto \widehat{\varphi_0(x+k)}$, when k runs through \mathbb{Z} . In general the family of functions $x \mapsto \varphi_0(x+k)$ ($k \in \mathbb{Z}$) is not orthogonal in $\mathbf{L}_2(\mathbb{R}, dx)$.

Proposition 4.10. *Let $W = |m_0|^2$, let*

$$(4.25) \quad h_{\varphi_0}(t) = \sum_{n \in \mathbb{Z}} |\widehat{\varphi_0}(t+n)|^2,$$

and let

$$(4.26) \quad (Rf)(t) = \frac{1}{N} \left(\sum_{k=0}^{N-1} (Wf)\left(\frac{t+k}{N}\right) \right).$$

Then

$$(4.27) \quad Rh_{\varphi_0} = h_{\varphi_0}.$$

Proof. We have

$$\begin{aligned}
(Rh_{\varphi_0})(t) &= \frac{1}{N} \sum_{k \in \mathbb{Z}_N} W\left(\frac{t+k}{N}\right) \sum_{n \in \mathbb{Z}} \left| \widehat{\varphi_0}\left(\frac{t+k+nN}{N}\right) \right|^2 \\
&= \sum_{k \in \mathbb{Z}_N} \sum_{n \in \mathbb{Z}} \left| \widehat{\varphi_0}\left(t + \overbrace{k+nN}^m\right) \right|^2 \\
&= \sum_{m \in \mathbb{Z}} \left| \widehat{\varphi_0}(t+m) \right|^2 \\
&= h_{\varphi_0}(t),
\end{aligned}$$

where we wrote \mathbb{Z}_N for the cyclic group $\mathbb{Z}/N\mathbb{Z}$, and we used the Euclidean algorithm on \mathbb{Z} , mod N , in the last step ($m = k + nN$). The first step used the scaling identity for φ_0 and $W = |m_0|^2$. \square

As an application of Proposition 4.10 we get the following results for wavelets on solenoids.

Corollary 4.11. *Let $W = |m_0|^2$ and h_{φ_0} be as in Proposition 4.10; then $M_n \stackrel{\text{def.}}{=} h_{\varphi_0}(\pi_n)$ is a $(\pi_n)_{n \in \mathbb{N}_0}$ -martingale, where $(\pi_n)_{n \in \mathbb{N}_0}$ denotes the $\text{Sol}_N(\mathbb{T})$ -Markov chain.*

Corollary 4.12. *Consider the wavelet filter m_0 with scaling function $\varphi_0 \in \mathbf{L}_2(\mathbb{R})$. Let h_{φ_0} be the corresponding harmonic function: $R_{m_0}h_{\varphi_0} = h_{\varphi_0}$, see (4.27). Then the level-0 isometry*

$$V_0 : f \in \mathbf{L}_2(\mathbb{T}, h_{\varphi_0}(t)dt) \mapsto (f(t)\widehat{\varphi_0}(t)) \in \mathbf{L}_2(\mathbb{R})$$

has the following explicit adjoint V_0^ computed on $\mathbf{L}_2(\mathbb{R})$:*

$$(V_0^*\gamma)(t) = \frac{1}{h_{\varphi_0}(t)} \sum_{n \in \mathbb{Z}} \gamma(t+n) \overline{\widehat{\varphi_0}(t+n)}, \quad t \in \mathbb{R}, \quad \gamma \in \mathbf{L}_2(\mathbb{R}, dx).$$

Remark 4.13. For functions k defined on $[0, 1]$ (or, equivalently, on \mathbb{R}/\mathbb{Z}) we introduce the Fourier coefficients

$$\widehat{k}(n) = \int_0^1 e^{-2\pi i n t} k(t) dt, \quad n \in \mathbb{Z}.$$

With $k = h_{\varphi_0}$ from (4.25) we can then compute the inner products $\int_{\mathbb{R}} \varphi_0(x+n) \overline{\varphi_0(x)} dx$ for $n \in \mathbb{Z}$. See the following proposition.

Proposition 4.14. *Let h_{φ_0} be the harmonic function associated to a scaling function $\varphi_0 \in \mathbf{L}_2(\mathbb{R}, dx)$. Then the following hold:*

(i)

$$\int_{\mathbb{R}} \varphi_0(x+n) \overline{\varphi_0(x)} dx = \widehat{h_{\varphi_0}}(n), \quad n \in \mathbb{Z}.$$

(ii) *The generating function*

$$\zeta \in \mathbb{C} \mapsto G_{\varphi_0}(\zeta) = \sum_{n \in \mathbb{Z}} \zeta^n \left(\int_{\mathbb{R}} \varphi_0(x+n) \overline{\varphi_0(x)} dx \right)$$

is an analytic extension of h_{φ_0} to an open neighborhood of \mathbb{T} .

(iii) *The scaling function φ_0 is compactly supported on \mathbb{R} if and only if G_{φ_0} is a polynomial.*

Proof. We need only to prove (i). The other two claims follow then easily. Using Parseval's equality in $\mathbf{L}_2(\mathbb{R})$ we have

$$\begin{aligned}
\int_{\mathbb{R}} \varphi_0(x+n) \overline{\varphi_0}(x) dx &= \int_{\mathbb{R}} e^{-2\pi i n t} |\widehat{\varphi_0}(t)|^2 dt \\
&= \sum_{m \in \mathbb{Z}} \int_0^1 e^{-2\pi i n t} |\widehat{\varphi_0}(t+m)|^2 dt \\
&\text{and using the dominated convergence theorem,} \\
&= \int_0^1 e^{-2\pi i n t} \left(\sum_{m \in \mathbb{Z}} |\widehat{\varphi_0}(t+m)|^2 \right) dt \\
&= \int_0^1 e^{-2\pi i n t} h_{\varphi_0}(t) dt \\
&= \widehat{h_{\varphi_0}}(n), \quad \forall n \in \mathbb{Z}.
\end{aligned}$$

□

Corollary 4.15. *Orthogonality of the family $\{\varphi_0(\cdot + n)\}_{n \in \mathbb{Z}}$ in $\mathbf{L}_2(\mathbb{R}, dx)$ is equivalent to the condition $h_{\varphi_0} \equiv 1$.*

In the next example we show that the Fejér kernels arise as h_{φ_0} for a family of scaling functions $\varphi_0 \in \mathbf{L}_2(\mathbb{R}, dx)$. We first recall that the Dirichlet kernel and Fejér kernels are defined respectively by

$$D_k(\zeta) = \sum_{j=-k}^k \zeta^j$$

and

$$F_k(\zeta) = \frac{\sum_{u=0}^k D_u(\zeta)}{k+1}.$$

Example 4.16. We take $\varphi_0(x) = \frac{1}{\sqrt{2m+1}} \chi_{[0, 2m+1]}(x)$, where $m \in \mathbb{N}$ is fixed. Then

$$\int_{\mathbb{R}} \varphi_0(x) \varphi_0(x-n) dx = \begin{cases} 0, & \text{if } |n| \geq 2m+1, \\ \frac{2m+1-n}{2m+1} & \text{for } n \in \{0, \dots, 2m\}. \end{cases}$$

Thus

$$(2m+1)h_{\varphi_0}(\zeta) = \zeta^{-2m} + 2\zeta^{1-2m} + \dots + (2m)\zeta^{-1} + (2m+1) + (2m)\zeta + \dots + 2\zeta^{2m-1} + \zeta^{2m},$$

which is the Fejér kernel F_{2m} .

4.3.3. *Realization using the solenoid.* We set $e_n(z) = z^n$, $n \in \mathbb{Z}$.

Theorem 4.17. *Let $z \in \mathbb{T}$. The function*

$$(4.28) \quad L\left(\frac{n}{N^k}\right) = (R^k(e_n h))(z)$$

is positive definite on $\mathbb{Z}[1/N]$, and there exists a positive finite measure $d\mu_z$ on $\text{Sol}_N(\mathbb{T})$ such that

$$(4.29) \quad L\left(\frac{n}{N^k}\right) = \int_{\text{Sol}_N(\mathbb{T})} \chi\left(\frac{n}{N^k}\right)(x) dP_z(x)$$

Proof: We first check that L is well defined. We have

$$\begin{aligned}
L\left(\frac{Nn}{N^{k+1}}\right) &= (R^{k+1}(e_{Nn}h))(z) \\
&= (R^k(Re_{nN}h))(z) \\
&= \left(R^k\left(\sum_{\substack{w \in \mathbb{T} \\ w^N = z}} |m_0(w)|^2 e_{nN}(w) h(w)\right)\right)(z) \\
&= \left(R^k\left(\sum_{\substack{w \in \mathbb{T} \\ w^N = z}} |m_0(w)|^2 e_n(z) h(w)\right)\right)(z) \quad (\text{since } e_{nN}(w) = e_n(z)) \\
&= \left(R^k\left(e_n \sum_{\substack{w \in \mathbb{T} \\ w^N = z}} |m_0(w)|^2 h(w)\right)\right)(z) \\
&= (R^k e_n R h)(z) \\
&= (R^k e_n h)(z) \\
&= L\left(\frac{n}{N^k}\right).
\end{aligned}$$

We now prove that L is positive definite on $\mathbb{Z}[1/N]$. Let $M \in \mathbb{N}$, $c_1, \dots, c_M \in \mathbb{C}$ and $\frac{n_1}{N^{k_1}}, \dots, \frac{n_M}{N^{k_M}} \in \mathbb{Z}[1/N]$. In view of the first part of the proof, we assume all the denominators equal, say to k . We have

$$\begin{aligned}
\sum_{u,v=1}^M \overline{c_u} c_v L\left(\frac{n_u}{N^k} - \frac{n_v}{N^k}\right) &= \sum_{u,v=1}^M \overline{c_u} c_v R^k((e_{n_u - n_v})h)(z) \\
&= \sum_{u,v=1}^M \overline{c_u} c_v R^k((e_{n_u - n_v})h)(z) \\
&= \sum_{u,v=1}^M \overline{c_u} c_v R^k((e_{n_u} \overline{e_{n_v}})h)(z) \\
&= (R(|g|^2 h))(z) \geq 0,
\end{aligned}$$

with $g = \sum_{u=1}^M \overline{c_u} e_u$.

The second claim comes from Bochner's theorem for compact groups. □

Let

$$\pi_0(z) = z_0,$$

and

$$U(\psi) = m_0(\pi_0(z))(\psi \circ \sigma_N)(z).$$

Proposition 4.18. *U is unitary and its adjoint is given by the formula*

$$(4.30) \quad U^* \psi = \frac{1}{m \circ \pi_1} \psi \circ \sigma_N^{-1}.$$

Proof: The results follow from the previous considerations; see also [42]. □

4.3.4. *Multiresolutions.* We set

$$\mathcal{L}_k = \text{closed linear span} \left\{ \chi \left(\frac{n}{N^k} \right), n \in \mathbb{Z} \right\}, \quad k \in \mathbb{Z}.$$

4.3.5. *Embedding the real line into the solenoid.* We define

$$(4.31) \quad \gamma_N(t) = \left(e^{2\pi i[t]}, e^{2\pi i[t/N]}, e^{2\pi i[t/N^2]}, \dots \right) \in \prod_{n=0}^{\infty} (\mathbb{R}/\mathbb{Z}),$$

where $[x]$ denotes the value of $x \in \mathbb{Z}$ modulo 1.

Lemma 4.19. *The map γ_N is one-to-one from \mathbb{R} into $\text{Sol}_N(\mathbb{T})$, meaning that*

$$\gamma_N(t) = (1, 1, 1, \dots) \iff t = 0.$$

Proof: See [42]. □

4.3.6. *Probability.* Let $h \geq 0$ be such that $Rh = h$, and assume that $\int_{\mathbb{T}} h(\lambda) d\lambda = 1$.

Proposition 4.20. *The distribution of the random variable*

$$\pi_k(\mathbf{z}) = z_k$$

is $|m^{(k)}(z)|^2 h(z) d\lambda$, where

$$(4.32) \quad m^{(k)}(z) = m_0(z) m_0(z^N) \cdots m_0(z^{N^{k-1}}).$$

Proof: We want to show that for every bounded measurable function f on \mathbb{T} we have

$$\int_{\mathbb{T}} f(e^{it}) d\lambda(e^{it}) = \int_{\text{Sol}_N(\mathbb{T})} f(\pi_k(\mathbf{z})) dP(\mathbf{z}).$$

We prove this equality for $f(z) = z^n$ (which we denoted by $e_n(z)$) with $n \in \mathbb{Z}$, that is

$$(4.33) \quad \int_{\mathbb{T}} e^{int} d\lambda(e^{it}) = \int_{\text{Sol}_N(\mathbb{T})} e_n(\pi(\mathbf{z})) dP(\mathbf{z}).$$

But recall that

$$\chi \left(\frac{n}{N^k} \right) (\mathbf{z}) = z_k^n = (\pi_k(\mathbf{z}))^n = e_n(\pi_k(\mathbf{z})).$$

Hence

$$\begin{aligned} \int_{\text{Sol}_N(\mathbb{T})} \chi \left(\frac{n}{N^k} \right) (\mathbf{z}) dP(\mathbf{z}) &= \int_{\mathbb{T}} R^k(e_n h)(e^{it}) d\lambda(e^{it}) \\ &= \int_{\mathbb{T}} (R^{*k}(1))(e^{it}) e_n(e^{it}) h(e^{it}) d\lambda(e^{it}) \\ &= \int_{\mathbb{T}} |m^{(k)}(e^{it})|^2 h(e^{it}) d\lambda(e^{it}), \end{aligned}$$

where we have used (4.24) to compute $R^{*k}1$. □

Remark 4.21. We note that

$$\begin{aligned} \int_{\mathbb{T}} |m^{(k)}(e^{it})|^2 h(e^{it}) d\lambda(e^{it}) &= \int_{\mathbb{T}} (R^{*k}(1))(e^{it}) h(e^{it}) d\lambda(e^{it}) \\ &= \int_{\mathbb{T}} (R^k(h))(e^{it}) d\lambda(e^{it}) \\ &= \int_{\mathbb{T}} h(e^{it}) d\lambda(e^{it}) = 1, \end{aligned}$$

as it should be.

4.3.7. The martingale property. As in the previous section h denotes a positive function such that $Rh = h$ and $\int_{\mathbb{T}} h(\lambda) d\lambda = 1$. Let $z, w \in \mathbb{T}$ be such that $w^N = z$, and set

$$P(z \mapsto w) = \frac{1}{N} |m(w)|^2 h(w).$$

The Markov property now reads

$$(4.34) \quad \sum_{\substack{w \in \mathbb{T} \\ w^N = z}} P(z \mapsto w) = 1.$$

The martingale property is now

$$R(\xi_{n+1}h) = \xi_n h,$$

and the following formulas hold:

$$\begin{aligned} \frac{R(fh)}{h} &\leq 1 \quad (\text{conditional expectation}) \\ \frac{R(\xi(z^n))h}{h} &= \xi \\ \sum_{w^N=z} \xi(w)h(w) &= \xi(z) \frac{Rh}{h} \\ M_{z_0 z_1} M_{z_1 z_2} \cdots &= R^k, \end{aligned}$$

with

$$M_{z_0 z_1} M_{z_1 z_2} = \frac{1}{N} \sum_{z_1^n = z_0} |m(z_1)|^2 \sum_{z_2^N = z_1} |m(z_1)|^2.$$

4.4. Fractal examples. We here consider X to be the set of numbers of the form

$$x = \sum_{n=1}^{\infty} \frac{b_n}{3^n}, \quad \text{where } b_n \in \{0, 2\}$$

and $\sigma(x) = 3x \pmod{1}$. In symbolic form we have

$$(b_1, b_2, b_3, \dots) \xrightarrow{\sigma} (b_2, b_3, b_4, \dots)$$

More generally, let $d \in \mathbb{N}$ and let $A \in \mathbb{Z}^{d \times d}$ with all eigenvalues of modulus strictly bigger than 1, and let $m < |\det A|$. Fix d residue classes b_1, \dots, b_d in $\mathbb{Z}^d / A\mathbb{Z}^d$. We set $Y = \{b_1, \dots, b_d\}$. We consider the set X of vectors in \mathbb{R}^d of the form

$$x = \sum_{n=1}^{\infty} A^{-n} c_n$$

where to make connections with homogeneous Markov chains (see (2.50) for the latter) we define

$$(4.35) \quad F_{b_j}(x) = A^{-1}(x + b_j), \quad j = 1, \dots, d$$

and $\sigma(x) = Ax$ modulo \mathbb{Z}^d .

Since Y is a finite set, a probability measure on Y is given by a finite number of positive numbers p_1, \dots, p_d adding up to 1, and the transfer operator is now given by

$$(4.36) \quad Rf = \sum_{j=1}^d p_j f \circ F_{b_j}.$$

Consider the set $\text{Prob}(X)$ of probabilities on X , and let $\nu, \mu \in \text{Prob}(X)$. Consider the Hausdorff distance between ν and μ :

$$d_H(\nu, \mu) = \sup \left\{ \int_X f(x) (d\nu(x) - d\mu(x)) \right\}$$

where the supremum is on the set of all Lipschitz functions:

$$|f(x) - f(y)| \leq \|x - y\|, \quad (\text{with } \|x - y\| \text{ being the usual distance in } \mathbb{R}^d).$$

Define a measure νR on X via

$$\int_X f(d(\nu R)) = \int_X (Rf) d\nu.$$

A theorem of Hutchinson (see [17, 37]) states that the map $\nu \mapsto \nu R$ is then strictly contractive. There exists $\alpha \in (0, 1)$ such that

$$d_H(R\nu, R\mu) \leq \alpha d_H(\nu, \mu).$$

Existence and uniqueness of a solution to the equation $\nu R = \nu$ follows from Banach fixed point theorem.

4.5. The Gauss operator. The present example is related to number theory and has links with information theory; see [15, 16, 55, 64]. We take $X = (0, 1)$ and $d\lambda(x) = dx$, and

$$(4.37) \quad \sigma(x) = \left\langle \frac{1}{x} \right\rangle,$$

where $\langle \cdot \rangle$ denotes the “fractional part”, defined as follows: If $x \in (\frac{1}{k+1}, \frac{1}{k})$ then $\sigma(x) = \frac{1}{x} - k$. We also define $\tau_k(x) = \frac{1}{x+k}$ with $k = 1, 2, \dots$. Note that

$$\sigma \circ \tau_k(x) = x, \quad k = 1, 2, \dots$$

The solenoid (see Definition 3.4) associated with X is described as follows:

$$(4.38) \quad \text{Sol}_\sigma(0, 1) = \left\{ (x_0, x_1, \dots) \in \prod_{n=0}^{\infty} (0, 1) \text{ such that } x_k = \left\langle \frac{1}{x_{k+1}} \right\rangle \right\}.$$

We thus obtain the continued fraction associated with x_0 .

$$\begin{aligned} x_0 &= k_1 + \frac{1}{\sigma(x_1)} \\ &= k_1 + \frac{1}{k_2 + \frac{1}{\sigma(x_2)}} \\ &= k_1 + \frac{1}{k_2 + \frac{1}{k_3 + \frac{1}{\ddots}}} \\ &\vdots \end{aligned}$$

Now the transfer operator is given by:

$$(4.39) \quad (Rf)(x) = \sum_{n=1}^{\infty} \frac{1}{(n+x)^2} f\left(\frac{1}{n+x}\right).$$

Proposition 4.22. *Let $h(x) = \frac{1}{\ln 2} \frac{1}{1+x}$. Then $\int_0^1 h(x) dx = 1$ and $\lambda R = \lambda$.*

Proof. We have

$$\begin{aligned} \lambda R = \lambda &\iff \int_0^1 \left(\sum_{n=1}^{\infty} \frac{1}{(n+x)^2} f\left(\frac{1}{n+x}\right) \right) dx = \int_0^1 f(x) dx \\ &\iff \sum_{n=1}^{\infty} \int_0^1 \frac{1}{(n+x)^2} f\left(\frac{1}{n+x}\right) dx = \int_0^1 f(x) dx. \end{aligned}$$

Note that the change of variable $y = \frac{1}{n+x}$ leads to

$$\int_0^1 \frac{1}{(n+x)^2} f\left(\frac{1}{n+x}\right) dx = \int_{\frac{1}{n+1}}^{\frac{1}{n}} f(y) dy,$$

and hence

$$\sum_{n=1}^{\infty} \int_0^1 \frac{1}{(n+x)^2} f\left(\frac{1}{n+x}\right) dx = \sum_{n=1}^{\infty} \int_{\frac{1}{n+1}}^{\frac{1}{n}} f(x) dx = \int_0^1 f(x) dx,$$

and the result follows. \square

REFERENCES

- [1] D. Alpay. *Algorithme de Schur, espaces à noyau reproduisant et théorie des systèmes*, volume 6 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 1998.
- [2] D. Alpay, P. Jorgensen, and I. Lewkowicz. Extending wavelet filters: infinite dimensions, the non-rational case, and indefinite inner product spaces. In *Excursions in harmonic analysis. Volume 2*, Appl. Numer. Harmon. Anal., pages 69–111. Birkhäuser/Springer, New York, 2013.

- [3] D. Alpay, P. Jorgensen, and I. Lewkowicz. Parametrizations of all wavelet filters: input-output and state-space. *Sampling Theory in Signal and Image Processing (STSIP)*, 12:159–188, 2013.
- [4] D. Alpay, P. Jorgensen, and I. Lewkowicz. Realizations of infinite products, Ruelle operators and wavelet filters. *J. Fourier Anal. Appl.*, 21(5):1034–1052, 2015.
- [5] C. Anantharaman and J. Renault. Amenable groupoids. In *Groupoids in analysis, geometry, and physics (Boulder, CO, 1999)*, volume 282 of *Contemp. Math.*, pages 35–46. Amer. Math. Soc., Providence, RI, 2001.
- [6] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.
- [7] L.W. Baggett. An abstract interpretation of the wavelet dimension function using group representations. *J. Funct. Anal.*, 173(1):1–20, 2000.
- [8] L.W. Baggett, V. Furst, K.D. Merrill, and J.A. Packer. Generalized filters, the low-pass condition, and connections to multiresolution analyses. *J. Funct. Anal.*, 257(9):2760–2779, 2009.
- [9] L.W. Baggett, V. Furst, K.D. Merrill, and J.A. Packer. Classification of generalized multiresolution analyses. *J. Funct. Anal.*, 258(12):4210–4228, 2010.
- [10] L.W. Baggett, K.D. Merrill, J.A. Packer, and A.B. Ramsay. Probability measures on solenoids corresponding to fractal wavelets. *Trans. Amer. Math. Soc.*, 364(5):2723–2748, 2012.
- [11] M. Bakonyi and T. Constantinescu. *Schur’s algorithm and several applications*, volume 261 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow, 1992.
- [12] V. Baladi. *Positive transfer operators and decay of correlations*, volume 16 of *Advanced Series in Nonlinear Dynamics*. World Scientific Publishing Co., Inc., River Edge, NJ, 2000.
- [13] G. Baumslag and D. Solitar. Some two-generator one-relator non-Hopfian groups. *Bull. Am. Math. Soc.*, 68:199–201, 1962.
- [14] S. Bezuglyi and P. Jorgensen. Representations of Cuntz-Krieger relations, dynamics on Bratteli diagrams, and path-space measures. In *Trends in harmonic analysis and its applications*, volume 650 of *Contemp. Math.*, pages 57–88. Amer. Math. Soc., Providence, RI, 2015.
- [15] N.M. Blachman. The information rate of a continued fraction (corresp.). *IEEE Trans. Inform. Theory*, 11:150–151, 1965.
- [16] N.M. Blachman. The continued fraction as an information source. *IEEE Trans. Inform. Theory*, 30(4):671–674, 1984.
- [17] O. Bratteli and P. Jorgensen. *Wavelets through a looking glass*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2002.
- [18] P. Brémaud. *Markov chains. Gibbs fields, Monte Carlo simulation, and queues*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999.
- [19] T. Constantinescu. *Schur parameters, factorization and dilation problems*, volume 82 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1996.
- [20] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)*, 39(1):1–49 (electronic), 2002.
- [21] J. Cuntz. Simple C^* -algebras generated by isometries. *Comm. Math. Phys.*, 57(2):173–185, 1977.
- [22] J. Cuntz and W. Krieger. Topological Markov chains with dicyclic dimension groups. *J. Reine Angew. Math.*, 320:44–51, 1980.
- [23] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
- [24] P. Diaconis and D. Freedman. Iterated random functions. *SIAM Rev.*, 41(1):45–76, 1999.
- [25] D. Dutkay and P. Jorgensen. Fourier duality for fractal measures with affine scales. *Math. Comp.*, 81(280):2253–2273, 2012.
- [26] D. Dutkay and P. Jorgensen. Spectral measures and Cuntz algebras. *Math. Comp.*, 81(280):2275–2301, 2012.
- [27] D. Dutkay and P. Jorgensen. The role of transfer operators and shifts in the study of fractals: encoding-models, analysis and geometry, commutative and non-commutative. In *Geometry and analysis of fractals*, volume 88 of *Springer Proc. Math. Stat.*, pages 65–95. Springer, Heidelberg, 2014.
- [28] D.E. Dutkay and P. Jorgensen. Wavelets on fractals. *Rev. Mat. Iberoam.*, 22(1):131–180, 2006.

- [29] H. Dym and V. Katsnelson. Contributions of Issai Schur to analysis. In *Studies in memory of Issai Schur (Chevaleret/Rehovot, 2000)*, volume 210 of *Progr. Math.*, pages xci–clxxxviii. Birkhäuser Boston, Boston, MA, 2003.
- [30] P. Erdős. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.*, 61:974–976, 1939.
- [31] P. Erdős. On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.*, 62:180–186, 1940.
- [32] R. Exel and J. Renault. Semigroups of local homeomorphisms and interaction groups. *Ergodic Theory Dynam. Systems*, 27(6):1737–1771, 2007.
- [33] P. Faure, M. Clerget, and F. Germain. *Opérateurs rationnels positifs*, volume 8 of *Méthodes Mathématiques de l’Informatique [Mathematical Methods of Information Science]*. Dunod, Paris, 1979. Application à l’hyperstabilité et aux processus aléatoires.
- [34] Ī. Ī. Gihman and A. V. Skorohod. *Controlled stochastic processes*. Springer-Verlag, New York-Heidelberg, 1979. Translated from the Russian by Samuel Kotz.
- [35] G. Hamel. Eine charakteristische Eigenschaft beschränkter analytischer Funktionen. *Math. Ann.*, 78(1):257–269, 1917.
- [36] O. Hernández-Lerma and J.B Lasserre. *Markov chains and invariant probabilities*, volume 211 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2003.
- [37] J.E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30(5):713–747, 1981.
- [38] M. Ionescu and P.S. Muhly. Groupoid methods in wavelet analysis. In *Group representations, ergodic theory, and mathematical physics: a tribute to George W. Mackey*, volume 449 of *Contemp. Math.*, pages 193–208. Amer. Math. Soc., Providence, RI, 2008.
- [39] M. Ionescu, P.S. Muhly, and V. Vega. Markov operators and C^* -algebras. *Houston J. Math.*, 38(3):775–798, 2012.
- [40] Y. Jiang and D. Ruelle. Analyticity of the susceptibility function for unimodal Markovian maps of the interval. *Nonlinearity*, 18(6):2447–2453, 2005.
- [41] Y. Jiang and Y.-L. Ye. Ruelle operator theorem for non-expansive systems. *Ergodic Theory Dynam. Systems*, 30(2):469–487, 2010.
- [42] P. Jorgensen. Ruelle operators: functions which are harmonic with respect to a transfer operator. *Mem. Amer. Math. Soc.*, 152(720):viii+60, 2001.
- [43] P. Jorgensen and A. Paolucci. States on the Cuntz algebras and p -adic random walks. *J. Aust. Math. Soc.*, 90(2):197–211, 2011.
- [44] P. Jorgensen and A. Paolucci. q -frames and Bessel functions. *Numer. Funct. Anal. Optim.*, 33(7-9):1063–1069, 2012.
- [45] S. Kakutani. On equivalence of infinite product measures. *Ann. of Math. (2)*, 49:214–224, 1948.
- [46] T. Katsura, P.S. Muhly, A. Sims, and M. Tomforde. Graph algebras, Exel-Laca algebras, and ultragraph algebras coincide up to Morita equivalence. *J. Reine Angew. Math.*, 640:135–165, 2010.
- [47] F. Latrémolière and J.A. Packer. Noncommutative solenoids and their projective modules. In *Commutative and noncommutative harmonic analysis and applications*, volume 603 of *Contemp. Math.*, pages 35–53. Amer. Math. Soc., Providence, RI, 2013.
- [48] F. Latrémolière and J.A. Packer. Explicit construction of equivalence bimodules between noncommutative solenoids. In *Trends in harmonic analysis and its applications*, volume 650 of *Contemp. Math.*, pages 111–140. Amer. Math. Soc., Providence, RI, 2015.
- [49] P.D. Lax and R.S. Phillips. Scattering theory for automorphic functions. *Bull. Amer. Math. Soc. (N.S.)*, 2(2):261–295, 1980.
- [50] P.D. Lax and R.S. Phillips. *Scattering theory*, volume 26 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, second edition, 1989. With appendices by Cathleen S. Morawetz and Georg Schmidt.
- [51] M. Loève. *Probability theory*. Third edition. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1963.
- [52] G.W. Mackey. *The scope and history of commutative and noncommutative harmonic analysis*, volume 5 of *History of Mathematics*. American Mathematical Society, Providence, RI; London Mathematical Society, London, 1992.
- [53] A.E. Marrero and P.S. Muhly. Groupoid and inverse semigroup presentations of ultragraph C^* -algebras. *Semigroup Forum*, 77(3):399–422, 2008.

- [54] J. Milnor. *Dynamics in one complex variable*. Friedr. Vieweg & Sohn, Braunschweig, 1999. Introductory lectures.
- [55] R. Nair. On the metrical theory of continued fractions. *Proc. Amer. Math. Soc.*, 120(4):1041–1046, 1994.
- [56] E. Nelson. *Topics in dynamics. I: Flows*. Mathematical Notes. Princeton University Press, Princeton, N.J., 1969.
- [57] E. Nelson. Quantum fields and Markoff fields. In *Partial differential equations (Proc. Sympos. Pure Math., Vol. XXIII, Univ. California, Berkeley, Calif., 1971)*, pages 413–420. Amer. Math. Soc., Providence, R.I., 1973.
- [58] J. Neveu. *Bases mathématiques du calcul des probabilités*. Masson et Cie, Éditeurs, Paris, 1964.
- [59] P. Niyogi, S. Smale, and S. Weinberger. A topological view of unsupervised learning from noisy data. *SIAM J. Comput.*, 40(3):646–663, 2011.
- [60] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.
- [61] J. Renault. AF equivalence relations and their cocycles. In *Operator algebras and mathematical physics (Constanța, 2001)*, pages 365–377. Theta, Bucharest, 2003.
- [62] D. Ruelle. Characteristic exponents and invariant manifolds in Hilbert space. *Ann. of Math. (2)*, 115(2):243–290, 1982.
- [63] D. Ruelle. *Thermodynamic formalism*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2004. The mathematical structures of equilibrium statistical mechanics.
- [64] C. Ryll-Nardzewski. On the ergodic theorems (ii) (ergodic theory of continued fractions). *Studia Mathematica*, 12(1):74–79, 1951.
- [65] I. Schur. Über die Potenzreihen, die im Innern des Einheitskreises beschränkten sind, I. *Journal für die Reine und Angewandte Mathematik*, 147:205–232, 1917. English translation in: I. Schur methods in operator theory and signal processing. (Operator theory: Advances and Applications OT 18 (1986), Birkhäuser Verlag), Basel.
- [66] S. Smale and D. Zhou. Online learning with Markov sampling. *Anal. Appl. (Singap.)*, 7(1):87–113, 2009.
- [67] B. Solomyak. On the random series $\sum \pm \lambda^n$ (an Erdős problem). *Ann. of Math. (2)*, 142(3):611–625, 1995.
- [68] H.S. Wall. *Analytic theory of continued fractions*. Van Nostrand, 1948.

(DA) DEPARTMENT OF MATHEMATICS, BEN-GURION UNIVERSITY OF THE NEGEV, BEER-SHEVA 84105 ISRAEL

E-mail address: dany@math.bgu.ac.il

(PJ) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF IOWA. IOWA CITY, IA 52242 USA

E-mail address: palle-jorgensen@uiowa.edu

(IL) DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING, BEN-GURION UNIVERSITY OF THE NEGEV, BEER-SHEVA 84105 ISRAEL

E-mail address: izchak@ee.bgu.ac.il